

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL



Ramo Automóvel - Uma Abordagem à Retenção

Henrique José Pires Marques

Mestrado em Matemática Aplicada à Economia e Gestão

Trabalho de Projeto orientado por:
Maria Fernanda Oliveira
Ricardo Martins Freire

Agradecimentos

Gostaria de agradecer à GENERALI e ao meu tutor Dr. Ricardo Martins Freire pela oportunidade de me proporcionarem um estágio na seguradora onde tive a possibilidade de aprender, desenvolver conhecimentos e por seguinte praticar todos os conceitos aprendidos.

Gostaria também de agradecer à professora Maria Fernanda Oliveira e à professora Teresa Alpuim, pelo seu tempo, ajuda e interesse ao longo do desenvolvimento deste relatório.

À minha equipa de trabalho Alexandra Mina e Evelina Correia pela sua disponibilidade ao longo de todo o tempo.

Queria ainda agradecer em especial aos meus pais, por todo o apoio prestado e a todas as pessoas que direta ou indiretamente estiveram relacionadas neste projeto.

E a todos os amigos, família e colegas de trabalho pelo apoio e motivação que me deram.

Resumo

Este projeto pretende através de conceitos e métodos aprendidos ao longo do mestrado em Matemática Aplicada à Economia e Gestão encontrar um modelo que explique o comportamento do Cliente, perante uma proposta de renovação do seguro automóvel.

Este modelo comportamental, puramente estatístico, serve de base para o cálculo da probabilidade de retenção de um determinado perfil de Cliente e foi posteriormente comparado com um modelo mais empírico desenvolvido pela seguradora Generali – Companhia de Seguros S.A. no âmbito das responsabilidades do Atuariado de Pricing Não Vida.

A preparação da base de dados e a sua modelização foi feita no software SAS, suportado pelo Excel para efeitos de gráficos e tabelas resumo. Foram utilizados dados referentes às propostas de renovação dos primeiros 8 meses de 2016. Previamente efetuou-se uma pré-modelização aos dados como forma de tratamento dos mesmos e evitarem-se assim erros classificatórios ou variáveis sem informação. Em última instância caso a correção de alguma observação não fosse possível ou estivesse incongruente com as restantes, optou-se por retirá-la deste estudo como forma de termos 100% de informação fidedigna e validada.

Para a criação deste modelo utilizou-se uma regressão logística múltipla, uma vez que a variável resposta pode tomar o valor de 1 (se o Cliente renovou contrato) ou 0 (caso o Cliente não tenha renovado contrato para a próxima renovação). Foram realizadas análises às possíveis variáveis explicativas com o intuito de tentar perceber se haveria alguma relação entre a variável e a probabilidade de retenção. Através de métodos estatísticos, criou-se um modelo de retenção que se revelou o mais ajustado e explicativo aos dados.

Por fim comparou-se o modelo puramente estatístico com o modelo existente na Generali, identificando-se as diferenças, analisando os resultados dos indicadores estatísticos e comparando a robustez de cada modelo.

Palavras-chave: Taxa de Retenção, Modelo de Regressão Logística, Seguro Automóvel

Abstract

This project intend through concepts and methods learned throughout the Masters in Mathematics Applied to Economics and Management to find a model that explain the behavior when a client receive a policy proposal for next renewal of motor insurance.

This behavioral model, purely statistical, serve as the bases for calculate the probability of retention of a certain customer profile and was later compared to a empirical model developed by the insurance company Generali within the scope of the responsibilities of the Non-Life Pricing Actuary department.

The preparation of the database and modelling was done in SAS software, supported by Excel for graphics and summary tables. The data referring to the proposals of renewal of the first 8 months of 2016 were used. Previously a pre-modeling was done to the data as a form of treatment of the same and avoiding thus classification errors or variables without information. Ultimately, if the correction of any observation was not possible or incongruent with the others, it was decided to delete it from the study as a way of having 100% reliable and validated information.

To create this model, a multiple logistic regression was used, since the response variable can take the value of 1 (if the customer renewed the policy) or 0 (if the customer didn't renew the policy for the next renewal). Were made analyses at explanatory variables in order to try to understand if there was any relationship between the variable and the probability of retention. By means of statistical methods, a retention model was created that proved to be the most accurate and explanatory of the data provided by Generali.

Finally, we compared the purely statistical model with the existing model used by Generali, identifying the differences, analyzing the results of the statistical indicators and comparing the robustness of each model.

Keywords: Retention Rate, Logistic Regression Model, Motor Insurance

Índice

Lista de Gráficos.....	3
Lista de Tabelas.....	4
Lista de Figuras.....	5
1. Introdução ao Mercado Segurador	6
1.1. Grupo GENERALI.....	6
1.2. Mercado Português.....	6
1.2.1. Ramo Vida.....	7
1.2.2. Ramo Não Vida.....	7
1.3. Outras Entidades	8
1.3.1. Autoridade de Supervisão de Seguros e Fundos de Pensões.....	8
1.3.2. Associação Portuguesa de Seguradores	8
2. Atividade Seguradora – Conceitos.....	9
3. Caracterização dos Ramos para a Seguradora.....	13
3.1. Seguros do Ramo Vida.....	13
3.1.1. Seguro em caso de Morte	13
3.1.2. Seguro em caso de Vida	13
3.1.3. Seguro Misto	13
3.2. Seguros do Ramo Não Vida	14
3.2.1. Seguro do Ramo Automóvel	14
3.2.2. Seguro do Ramo de Transportes	14
3.2.3. Seguros do Ramo Patrimonial.....	14
3.2.4. Seguros do Ramo de Responsabilidade Civil.....	15
3.2.5. Seguro do Ramo de Saúde	15
3.2.6. Seguros de Acidentes de Trabalho	15
3.2.7. Seguros de Acidentes Pessoais.....	15
4. Ramo Automóvel	16
4.1. Categoria de Veículos	16
4.2. Seguro de Responsabilidade Civil.....	17
4.3. Seguro de Danos Próprios	18
4.4. Outros Danos.....	18
4.5. Bónus-Malus	19
4.6. Fundo de Garantia Automóvel	20
4.7. Regularização de Sinistros	20

5.	Análise Estatística	21
5.1.	Regressão	21
5.1.1.	Regressão Logística.....	21
5.2.	Indicadores de Comparação	23
5.2.1.	Matriz de Confusão	23
5.2.2.	Curva ROC e Área Abaixo da Curva	24
5.2.3.	Critério de Informação de Akaike e Bayesiano.....	27
5.3.	Multicolinearidade	28
5.4.	Tipos de Seleção Possíveis.....	28
5.4.1.	Método de Seleção Regressiva.....	29
6.	Construção do Modelo	30
6.1.	Escolha de Observações	30
6.2.	Preparação dos Dados	31
6.3.	Análise Univariada.....	32
6.3.1.	Transformação de dados.....	32
6.3.2.	Análise Descritiva	33
6.4.	Seleção de Variáveis	50
6.5.	Diagnóstico do Modelo	59
6.5.1.	Matriz de Confusão e Curva de ROC.....	60
6.5.2.	Indicadores do modelo construído	62
7.	Modelo Generali.....	63
8.	Conclusões	67
9.	Referências	70
9.1.	Referências Bibliográficas	70
9.2.	Referências Sitográficas	70
10.	Anexos.....	71

Lista de Gráficos

Gráfico 5.1 - Sensibilidade e Especificidade.....	25
Gráfico 6.1 - Taxa de Retenção por Tipo de Veículo	32
Gráfico 6.2 - Taxa de Retenção por Tipo de Veículo Base 100.....	33
Gráfico 6.3 - Taxa de Retenção por Zona Geográfica	34
Gráfico 6.4 - Taxa de Retenção por Tipo de Cliente.....	34
Gráfico 6.5 - Taxa de Retenção por Grupo de Agentes	35
Gráfico 6.6 - Taxa de Retenção por Mês de Renovação	36
Gráfico 6.7 - Taxa de Retenção por Tipo de Veículo	36
Gráfico 6.8 - Taxa de Retenção por Existência da Cobertura de Quebra Isolada de Vidros	37
Gráfico 6.9 - Taxa de Retenção por existência da cobertura de Danos Próprios	38
Gráfico 6.10 - Taxa de Retenção por franquia de Danos Próprios.....	38
Gráfico 6.11 - Taxa de Retenção por Letra de Bónus-Malus.....	39
Gráfico 6.12 - Taxa de Retenção por prêmio pago na anuidade de 2015	40
Gráfico 6.13 - Taxa de Retenção por Prêmio proposto	40
Gráfico 6.14 - Taxa de Retenção por Variação do Prêmio em Percentagem	41
Gráfico 6.15 - Taxa de Retenção por Frequência de Pagamento	42
Gráfico 6.16 - Taxa de Retenção por Marca do Veículo.....	43
Gráfico 6.17 - Taxa de Retenção por Tipo de Combustível.....	43
Gráfico 6.18 - Taxa de Retenção por Rácio Peso-Potência.....	44
Gráfico 6.19 - Taxa de Retenção por Distrito	45
Gráfico 6.20 - Taxa de Retenção por Valor do Veículo.....	45
Gráfico 6.21 - Taxa de Retenção por Idade do Condutor	46
Gráfico 6.22 - Taxa de Retenção por Idade do Segurado	47
Gráfico 6.23 - Taxa de Retenção por Idade do Veículo	48
Gráfico 6.24 - Taxa de Retenção por Antiguidade da Carta de Condução	49
Gráfico 6.25 - Taxa de Retenção por Antiguidade da Apólice	49
Gráfico 6.26 - Taxa de Retenção por Retenção Média do Agente	50
Gráfico 6.27 - Sensibilidade e Especificidade do modelo criado.....	61
Gráfico 6.28 - Curva ROC do modelo criado	62
Gráfico 7.1 - Sensibilidade e Especificidade do Modelo Generali	65
Gráfico 7.2 - Curva ROC do modelo da Generali.....	66

Lista de Tabelas

Tabela 1.1 - Ramo Vida	7
Tabela 1.2 - Ramo Não Vida.....	7
Tabela 5.1 – Diferentes pontos de corte	25
Tabela 6.1 - Taxa de Retenção Por Tipo de Veículo.....	32
Tabela 6.2 - Multicolinearidade das variáveis - Fase 1	51
Tabela 6.3 - Multicolinearidade das variáveis - Fase 2	51
Tabela 6.4 - Multicolinearidade das variáveis - Fase 3	52
Tabela 6.5 - Exemplo de Variável Dummy	52
Tabela 6.6 - Variáveis a eliminar do modelo	53
Tabela 6.7 - Variáveis selecionadas para o modelo	55
Tabela 6.8 - Diferentes pontos de corte para o modelo criado	60
Tabela 6.9 - Matriz de confusão do modelo criado para o seu ponto de corte perfeito.....	61
Tabela 6.10 - Matriz de confusão do modelo criado para o ponto de corte de 78.38%	61
Tabela 7.1 – Output Emblem	63
Tabela 7.2 - Diferentes pontos de corte para o modelo Generali	64
Tabela 7.3 – Matriz de confusão para o ponto de corte de 77.9% referente ao modelo Generali	65
Tabela 7.4 – Matriz de confusão para o ponto de corte perfeito do modelo Generali	66

Lista de Figuras

Figura 2.1 - Detalhe de uma Apólice	11
Figura 5.1 - Matriz de Confusão	23
Figura 5.2 - Curva ROC	26
Figura 8.1 - Margem e Retenção.....	67

1. Introdução ao Mercado Segurador

1.1. Grupo GENERALI

O Grupo Generali foi criado no dia 26 de dezembro de 1831 com o nome de *Assicurazioni Generali Austro-Italiche* na cidade de Trieste, com um capital de 2 milhões de fiorinos Austriacos. Apesar de Trieste ser um dos principais portos marítimos regionais, tinha uma grande particularidade – os ventos bora. Estes ventos são caracterizados por terem uma forte potência que faz com que o mar fique bastante agitado, e por sinal uma sinistralidade elevada. Desta forma surgiu a necessidade de se criar uma companhia sólida com o intuito de se tornar numa das mais importantes do mundo. Rapidamente se começou a expandir, sendo que, 10 anos mais tarde, já estava presente nos principais centros Italianos e em alguns dos principais portos marítimos europeus. Em 1848, o grupo decidiu mudar o seu nome para *Assicurazioni Generali* e adotou o leão de S. Marcos como seu símbolo. Após o fim da Primeira Guerra Mundial (1918), Trieste foi anexada a Itália, obtendo o grupo o estatuto de nacionalidade Italiana.

À data de hoje, o grupo Generali está presente em mais de 50 países e 3 Continentes, possuindo mais de 74 000 colaboradores, atingindo no ano de 2016 um resultado operacional de 4.8 mil milhões de euros, graças também aos seus 150 mil agentes exclusivos. É o terceiro maior grupo Europeu do Setor Segurador. Posiciona-se também, entre os líderes da indústria em termos de responsabilidade e desempenho.

Em Portugal, o grupo começou a operar em 1942 e está presente em todo o país, incluindo ilhas através de 14 delegações, possuindo cerca de 350 trabalhadores. Em 2016 atingiu um volume de Prémios Brutos Emitidos de 190 Milhões de Euros.

1.2. Mercado Português

Embora pequeno, o mercado Português é bastante competitivo, pelo que cabe às seguradoras encontrarem formas inovadoras para maximizarem o seu lucro. A tabela seguinte mostra como está distribuído o mercado português pelas 15 principais seguradoras nos diversos ramos. O valor dos prémios brutos está em milhares de euros.

1.2.1. Ramo Vida

Tabela 1.1 - Ramo Vida

Lugar	Companhia de Seguros	Prémios Brutos 2015	Prémios Brutos 2016	Quota de Mercado	Δ (%)
1	Fidelidade	2.723.596 €	2.369.754 €	35,6%	-13,0%
2	Ocidental Vida	1.533.391 €	1.408.265 €	21,2%	-8,2%
3	BPI Vida	1.751.140 €	488.372 €	7,3%	-72,1%
4	Santander Totta Seguros Vida	142.683 €	430.901 €	6,5%	202,0%
5	Allianz Portugal	282.886 €	323.491 €	4,9%	14,4%
6	Zurich Vida	138.851 €	175.400 €	2,6%	26,3%
7	CA Vida	344.648 €	168.379 €	2,5%	-51,1%
8	Lusitania Vida	138.572 €	162.190 €	2,4%	17,0%
9	GNB Seguros Vida	464.467 €	153.073 €	2,3%	-67,0%
10	Eurovida	167.201 €	144.060 €	2,2%	-13,8%
11	Bankinter Seguros de Vida	- €	140.312 €	2,1%	0,0%
12	Ageas Portugal Vida	128.244 €	116.825 €	1,8%	-8,9%
13	Groupama Vida	95.650 €	100.207 €	1,5%	4,8%
14	Generali Vida	66.492 €	61.851 €	0,9%	-7,0%
15	Aegon Santander Vida	53.735 €	57.510 €	0,9%	7,0%

Num ano (2016) em que o mercado Português teve uma diminuição de 22.7% em relação ao volume de prémios no ramo vida, a seguradora Generali ocupa o 14º lugar no ranking nacional, atingindo um volume de prémios no último ano de aproximadamente 62 Milhões de Euros. Neste ramo, as seguradoras líderes (Fidelidade e Ocidental Vida) possuem mais de 56% da quota de mercado.

1.2.2. Ramo Não Vida

Tabela 1.2 - Ramo Não Vida

Lugar	Companhia de Seguros	Prémios Brutos 2015	Prémios Brutos 2016	Quota de Mercado	Δ (%)
1	Fidelidade	997.077 €	1.082.512 €	26,5%	8,6%
2	Allianz Portugal	338.923 €	374.601 €	9,2%	10,5%
3	Tranquilidade	320.292 €	341.062 €	8,4%	6,5%
4	Ocidental	259.362 €	285.492 €	7,0%	10,1%
5	Liberty Seguros	250.854 €	284.787 €	7,0%	13,5%
6	Açoreana	274.558 €	282.949 €	6,9%	3,1%
7	Ageas Portugal Não Vida	247.987 €	259.155 €	6,4%	4,5%
8	Zurich Insurance	238.317 €	226.197 €	5,5%	-5,1%
9	Lusitania Seguros	184.801 €	185.093 €	4,5%	0,2%
10	Generali - Companhia de Seguros	129.586 €	128.543 €	3,2%	-0,8%
11	Mapfre Seguros Gerais	99.997 €	96.549 €	2,4%	-3,4%
12	CA Seguros	87.075 €	95.890 €	2,4%	10,1%
13	Victoria Seguros	74.121 €	77.559 €	1,9%	4,6%
14	GNB Seguros	69.974 €	71.602 €	1,8%	2,3%
15	Via Directa	44.083 €	45.535 €	1,1%	3,3%

Já no ramo não vida, o número de seguradoras a lutarem pela liderança é maior, sendo que as 10 maiores seguradoras detêm mais de 84 % da quota de mercado.

A Generali está nesse top 10, tendo uma quota de mercado de 3.4% no ano de 2016, valor este ligeiramente inferior a período homologado. Este ramo tem um volume de prémios de 4 mil milhões de euros, tendo o mercado sofrido um aumento de 6% em relação a 2015.

1.3. Outras Entidades

1.3.1. Autoridade de Supervisão de Seguros e Fundos de Pensões

A Autoridade de Supervisão de Seguros e Fundos de Pensões (ASF) é a entidade responsável pela regulação e supervisão das seguradoras, resseguradoras, fundos de pensões e mediação de Seguros. Serve para assegurar o bom funcionamento do mercado segurador e fundos de pensões, com o intuito de proteger os tomadores de seguros e segurados. Esta entidade tem poder sobre todas as seguradoras, podendo assim assegurar a solidez financeira das mesmas através de regras e padrões definidos por ela.

Estão inscritas na ASF 34 seguradoras dos diversos ramos.

1.3.2. Associação Portuguesa de Seguradores

A Associação Portuguesa de Seguradores (APS) tem como objetivo defender e promover os interesses das seguradoras. Representa e defende os interesses dos associados e divulga as suas posições. Tem também como função a criação de cursos para associados, a realização de estudos ou de ações que sejam do interesse do associado ou de qualquer entidade seguradora.

2. Atividade Seguradora – Conceitos

Um Contrato de Seguro é um contrato celebrado entre uma companhia de Seguros e uma pessoa (singular ou coletiva), no qual a companhia se compromete a pagar uma indemnização a um indivíduo (segurado) que sofra um sinistro, mediante um prémio de seguro. Como tal, assinam um contrato de seguro que requer um risco. Na perspetiva de uma seguradora, o prémio de seguro tem de ser superior ao que esta espera vir a pagar ao cliente em caso de sinistro. Por seu turno, da perspetiva do cliente, este quer pagar o mínimo possível. Por isso, o prémio não pode ser demasiado elevado nem baixo.

Nem sempre as seguradoras querem ou podem assumir o risco na sua totalidade, pelo que surgiram contratos de co-seguro e resseguro. Um contrato de co-seguro consiste na partilha do risco com determinadas congéneres, de forma a que o risco fique dividido. A seguradora líder é a que emite o contrato de co-seguro, e este contém todas as seguradoras que partilham o risco da apólice bem como a sua quota-parte. O prémio é dividido de forma percentual, assim como o custo de sinistros.

Um contrato de resseguro é um contrato, mediante o qual, a seguradora transfere para outrem todo, ou parte do risco da sua responsabilidade. Tem como principal objetivo reduzir a importância de uma determinada perda. A grande diferença entre um contrato de resseguro e um contrato de co-seguro está na repartição do risco. Enquanto que nos contratos de resseguro o segurado permanece alheio a tal repartição, não lhe sendo reconhecido o direito de se pronunciar sobre qualquer ressegurador, o mesmo não acontece no co-seguro .

Uma vez que um contrato de seguro pode ser dividido por várias companhias de seguro, é importante definir diversos conceitos acerca da atividade Seguradora:

- Risco – é uma combinação entre a probabilidade de ocorrência de um determinado sinistro e os prejuízos daí resultantes;
- Apólice de Seguro – conjunto de condições identificadas e na qual é formalizado um contrato de Seguro;
- Sinistro – é o evento que desencadeia o segurado a acionar uma cobertura de risco prevista no contrato de seguro;
- Retenção – Ação que ocorre quando o segurado recebe uma proposta de renovação do seu contrato de seguro por parte da companhia e a aceita.

No que diz respeito a intervenientes num contrato de seguro, destaca-se:

- Seguradora – Entidade responsável pela criação de um contrato de seguro. Esta recebe o devido prémio de seguro e paga o respetivo sinistro;
- Tomador do Seguro – Pessoa ou entidade que contrata um seguro, sendo esta apenas caracterizada pelo pagamento de prémios;
- Segurado – Pessoa ou a entidade titular do interesse do seguro. É sobre esta pessoa que é calculado o prémio de risco;
- Mediador – É a entidade que se dedica à intermediação de seguros em nome de uma seguradora;
- Agente – Pessoa responsável pela venda de contratos de seguros.

Um contrato de seguro não é apenas caracterizado por intervenientes. Também é constituído por diversos elementos formais, nomeadamente:

- Condições gerais da apólice – Conjunto de condições contratuais, definindo os princípios gerais do contrato, independentemente do tipo de risco;
- Condições particulares – Conjunto de condições extra que não figuram nas condições gerais da apólice. Estas acrescentam cláusulas particulares de cada indivíduo. É exemplo a duração do contrato ou o capital seguro (ver página seguinte);
- Certificado de Seguro – Documentos que comprovam a existência de um contrato de seguro;
- Proposta de alteração – Proposta que permite ao segurado realizar alterações na sua apólice. Um exemplo é a alteração de capital seguro.

Para além destes termos, um contrato de seguro é caracterizado por ter diversos conceitos acerca do prazo:

- Data de comercialização – Data em que o segurado inicia o seu acordo com a companhia de seguros;
- Início da vigência – Data a partir da qual a apólice está em risco;
- Fim da vigência – Data a partir da qual a apólice deixa de estar em risco;
- Prazo – Período combinado para a sua Vigência;
- Duração do contrato – A duração de um contrato de seguro é o tempo em que o contrato está em vigor. Este pode ser de dois tipos:
 - Duração certa ou determinada – Um contrato tem duração certa ou determinada quando a apólice respetiva é temporária, ou seja, tem um prazo limitado não sendo posteriormente renovada. Esta apólice inicia no dia de início da vigência e termina no dia do fim da Vigência;
 - Duração de um ano com renovação anual – Este tipo de duração também é chamado de Anos e Seguintes. A respetiva apólice começa no início da vigência e acaba no fim da mesma, sendo que é renovada automaticamente todos os anos até à sua anulação. É o tipo mais comum existente numa seguradora.
- Data de anulação da apólice – Data em que o segurado ou seguradora anulou a apólice;
- Data de ocorrência de um sinistro – Data em que o segurado esteve envolvido num sinistro;
- Data de aviso de um sinistro – Data em que o segurado avisa a companhia de seguros da ocorrência de um sinistro. Muitas vezes não coincide com a data de ocorrência do sinistro.

Apresentamos de seguida um exemplo ilustrativo das datas e durações de Contrato:

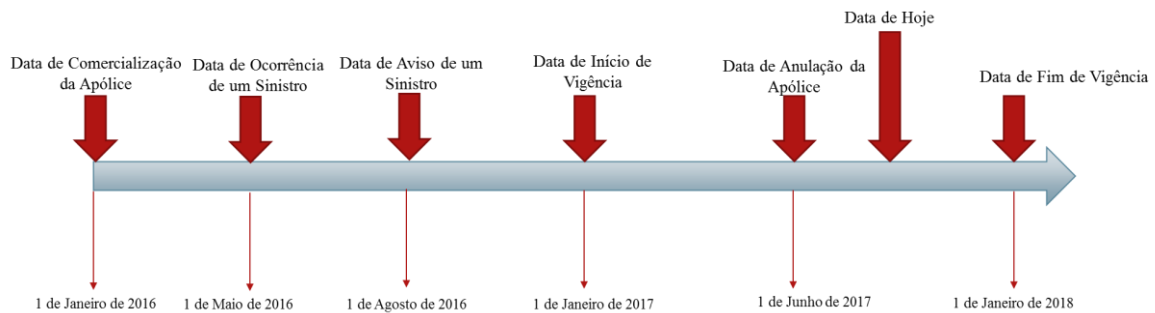


Figura 2.1 - Detalhe de uma Apólice

A figura anterior ilustra uma apólice que comercializou no dia 1 de Janeiro de 2016, com duração de um ano, que foi renovada automaticamente ao fim do mesmo. À data de hoje a apólice já está anulada por decisão do segurado sendo que o início de vigência é 1 de Janeiro de 2017 e deveria terminar no fim de vigência (1 de Janeiro de 2018). A apólice teve um sinistro no dia 1 de Maio de 2016, que foi avisado à companhia no dia 1 de Agosto de 2016, sendo este um caso em que a data de ocorrência e aviso do sinistro não coincidem.

Um contrato de seguro também é caracterizado por ter um capital seguro. Este valor corresponde ao máximo que a companhia está disponível a pagar aquando de um sinistro.

Aquando de um sinistro, é também importante definir alguns conceitos:

- Beneficiário – A pessoa ou pessoas que receberão qualquer valor resultante de um sinistro no qual o segurado esteve envolvido. Em caso de não ser atribuída culpa ao segurado, é a seguradora que paga esse montante;
- Terceiro – Aquele que em consequência de um sinistro, está envolvido no mesmo, ainda que não diretamente;
- Dano Corporal – Prejuízo resultante de uma lesão de saúde física ou mental ocorridos num sinistro;
- Dano Material – Prejuízo resultante de uma lesão de partes moveis, imoveis ou materiais;
- Franquia – Valor da regularização do sinistro nos termos do contrato de seguro que não fica a cargo da Seguradora, ou seja, é o valor que o segurado irá sempre pagar em consequência de um sinistro;
- Valor em Novo – Preço de Venda ao publico de um veículo seguro pela companhia. Este conceito só é praticamente usado no ramo Automóvel;
- Valor de Substituição – Valor comercial médio cotado no mercado de veículos usados, para a aquisição por parte do Tomador de Seguro, no momento do sinistro, de um veículo com as mesmas características e estado de conservação;
- Valor Venal – Valor comercial médio cotado no mercado de veículos usados, para venda por parte do tomador de Seguro, no momento do sinistro, de um veículo com as mesmas características e estado de conservação;
- Perda Total – Desaparecimento do veículo seguro ou destruição do mesmo quando se verifique uma das seguintes opções:

- A reparação não seja materialmente possível ou tecnicamente aconselhada;
- A reparação seja possível, mas o seu custo seja maior que o valor do veículo, valor este determinado pela tabela de desvalorização de veículos.

Um conceito também importante em todo o tipo de seguros é o conceito de prémio:

- Prémio de risco – Prémio que o cliente irá pagar à companhia para cobrir o risco da sua apólice. Muitas vezes é calculado através de modelos estatísticos criados pela empresa através do histórico de outros clientes;
- Prémio comercial – O valor do prémio comercial é igual ao prémio de risco líquido de descontos/agravamentos que a companhia fornece ao cliente acrescido de custo de aquisição, administração do contrato, gestão e de cobrança;
- Prémio bruto – O valor do prémio bruto é igual ao prémio comercial acrescido de cargas relacionadas com a emissão da apólice, como por exemplo fracionamento, custo de apólices, certificados de seguro e atas adicionais;
- Prémio total – O valor do prémio total é o valor que o cliente irá pagar à seguradora pelo seu contrato de seguro. É igual ao prémio bruto acrescido de valores que a seguradora tem de pagar a entidades externas, ou seja, cargas fiscais e para fiscais.

3. Caracterização dos Ramos para a Seguradora

Na atividade seguradora, assim como na Generali, os seguros são caracterizados por estarem atribuídos a um de dois tipos de seguro: os seguros do ramo vida e os seguros do ramo não vida.

3.1. Seguros do Ramo Vida

Os seguros do ramo vida são seguros que têm como cobertura principal segurar o risco de morte ou de vida de um indivíduo ou conjunto de indivíduos.

Temos então os três principais tipos de seguros disponíveis:

3.1.1. Seguro em caso de Morte

Caso uma pessoa queira segurar o risco de morte durante um determinado período de tempo, pode fazer o chamado “seguro em caso de morte” e a seguradora será responsável por pagar uma indemnização em caso de morte do segurado aos seus beneficiários. É o tipo de seguros mais comuns no ramo Vida.

3.1.2. Seguro em caso de Vida

De igual forma, um indivíduo pode querer segurar a sua vida. Desta forma, acorda com a seguradora que caso sobreviva durante um determinado período, a seguradora terá de lhe pagar um Capital. Caso ela morra durante um determinado período, a seguradora fica com a totalidade do valor do prémio que recebeu e não tem de pagar nada ao segurado.

3.1.3. Seguro Misto

Por fim, existe, também, a opção de um contrato misto, ou também chamado contrato em caso de vida e de morte. Aqui o segurado compra o seguro que cubra a sua morte nos primeiros x anos, e caso sobreviva a esses anos recebe outro montante.

3.2. Seguros do Ramo Não Vida

Os seguros do Ramo Não Vida são seguros que têm como objetivo segurar um bem. É neste tipo de seguro que estão praticamente todos os seguros que conhecemos. Sendo assim, a companhia está dividida em diversos sub-ramos:

3.2.1. Seguro do Ramo Automóvel

Este seguro cobre veículos ou articulados que circulem na via rodoviária. Tem uma parte de responsabilidade civil obrigatória, não podendo nenhum veículo circular sem um seguro com esta particularidade. Existe também uma componente extra. Dentro deste tipo de seguros estão englobados os automóveis, camiões, autocarros, caravanas, motociclos e ciclo-motores, entre outros.

3.2.2. Seguro do Ramo de Transportes

Neste tipo de seguros estão englobados veículos que circulam no ar, em água ou em terra. Tem também a componente obrigatória de responsabilidade civil. É também possível segurar a mercadoria em qualquer meio de transporte.

3.2.3. Seguros do Ramo Patrimonial

Este subdivide-se em três variantes importantes:

3.2.3.1. Seguro Multirrisco de Habitação:

É um seguro obrigatório em Portugal, sendo que o proprietário tem a obrigação de cobrir o risco em caso de um incêndio na sua fração, ou nas partes comuns do edifício. Têm também a opção de segurar o recheio da habitação. Existe também a possibilidade de subscrever garantias complementares como é o caso dos fenómenos sísmicos

3.2.3.2. Seguro Multirrisco Industrial e de Comerciantes:

Os seguros multirrisco industrial são seguros que têm como objetivo segurar um espaço resultante do sector industrial. Por outro lado também se pode segurar a atividade resultante do comércio. Aqui o segurado pode cobrir tanto o edifício como o seu conteúdo.

3.2.4. Seguros do Ramo de Responsabilidade Civil

Neste tipo de seguro, a seguradora cobre o risco do segurado ter de vir a indemnizar um terceiro por um dano causado por ele.

3.2.5. Seguro do Ramo de Saúde

Este é caracterizado por cobrir cuidados de saúde por parte de clientes, como idas ao médico, cirurgias, fisioterapia, entre outros. Possuem muitas vezes um capital máximo a utilizar por parte do segurado, bem como uma franquia para o uso do mesmo.

3.2.6. Seguros de Acidentes de Trabalho

O seguro de Acidentes de Trabalho é um seguro que tem como característica proteger o segurado de qualquer acidente que tenha, tanto no local de trabalho como no trajeto casa-trabalho-casa. É um seguro obrigatório para todos os trabalhadores por conta doutrem. Assim, caso o segurado tenha algum dano decorrente de um acidente nestes parâmetros, está salvaguardado. Este seguro garante despesas de tratamento e indemnizações salariais (vulgarmente baixas). Este tipo de seguro é uma particularidade do mercado segurador português, não existindo em muitos países.

3.2.7. Seguros de Acidentes Pessoais

O seguro de Acidentes Pessoais é um seguro que garante um pagamento de indemnizações previamente definidas pela ocorrência de um acidente, na sua vida particular. É um seguro cada vez mais presente no mercado segurador, ficando o segurado coberto 24 horas por dia de um eventual azar. É neste tipo de seguros que se encontram os seguros de viagem.

4. Ramo Automóvel

4.1. Categoria de Veículos

De acordo com Código da Estrada, os veículos podem ser classificados como:

- Automóveis – Veículos com motor e que não necessitam de carris para se deslocar. Estes dividem-se em dois tipos:
 - Ligeiros – São veículo caracterizados por ter um peso bruto que não excede os 3500 quilogramas e que tem uma lotação máxima de 9 pessoas incluindo o condutor;
 - Pesados – São veículos caracterizados por ter um peso bruto maior que 3500 quilogramas ou com lotação superior a 9 lugares;

Estas duas classes de veículos podem, ainda, ser divididas consoante o seu tipo de utilização:

- Passageiros – Veículos que se destinam ao transporte de pessoas;
 - Mercadorias – Veículos de transporte de carga;
- Motociclos, ciclomotores, triciclos e quadriciclos – Motociclos e triciclos são veículos com menos de 4 rodas cujo motor tem no máximo 50cm³. Já os ciclomotores são veículos de duas ou três rodas que não excede os 45 quilómetros por hora e os quadriciclos têm as mesmas características dos ciclomotores, mas possuem 4 rodas;
- Veículos agrícolas – Veículos que se destinam a trabalhos agrícolas, sendo eles tratores agrícolas, motocultivador, máquinas agrícolas ou florestais. Apenas máquinas agrícolas ou florestais podem circular na estrada;
- Outros veículos a motor – São Máquinas industriais que excecionalmente circulam na via pública e veículos a motor que circulam sobre carris;
- Reboque – Veículo sem motor que circula atrelado ao veículo que proporciona a força para andar;
- Velocípedes – Veículo de uma ou duas rodas que se move graças ao esforço do condutor por via de pedais.

Quando um cliente quer segurar um veículo, a seguradora propõe a contratualização de diversos tipos de contrato. Há a opção de contratos de Seguros de Responsabilidade Civil, de Danos Próprios ou de Outros danos.

4.2. Seguro de Responsabilidade Civil

O Seguro automóvel é um seguro obrigatório para todos os que pretendem circular com um veículo. Existe uma parte obrigatória (cobertura de Responsabilidade Civil) que todos os indivíduos têm de ter e que tem como principal objetivo proteger a outra parte de um possível sinistro.

Dentro deste tipo de seguro, há duas espécies de contratos disponíveis:

- Seguro Obrigatório: É o seguro celebrado nos termos da lei. Atualmente o mínimo exigido por lei é que o segurado tenha um capital de 7.290.000 €, sendo que está dividido em 6.070.000 € para Danos Corporais e 1.220.000 € para Danos Materiais. Até junho de 2017, o capital obrigatório era de 6.000.000 €;
- Seguro Facultativo: É um contrato celebrado por valores superiores aos mínimos exigidos por lei, ou por veículos que não são sujeitos à parte obrigatória. Assim é possível o aumento do capital mínimo obrigatório de 7.290.000 € para 50.000.000 €.

Quando os veículos têm como finalidade o transporte coletivo de passageiros (caso dos táxis, autocarros,...) , o mínimo obrigatório é o dobro, ou seja, 14.580.000 €.

Sendo assim, este seguro cobre:

- Acidentes ocorridos em Território Nacional, a obrigação da companhia de seguros indemnizar até ao capital seguro pelo cliente;
- Acidentes ocorridos fora do Território Nacional e abrangidos pelo Acordo do Espaço Económico Português, a obrigação de indemnizar até ao capital mínimo exigido pela legislação do país ocorrente ou pela legislação Portuguesa quando este for superior;
- Danos corporais e materiais provocados a terceiros pelo veículo segurado.

No entanto existem diversas exclusões possíveis, ou seja, este contrato de seguro não cobre:

- Sinistros ocorridos num país que não tenha aderido ao serviço nacional de seguros;
- Danos corporais sofridos pelo condutor do veículo seguro se for o responsável pelo acidente, bem como os danos dele decorrentes;
- Danos materiais sofridos pelo veículo seguro se este sofrer um sinistro e for de sua responsabilidade.

4.3. Seguro de Danos Próprios

Quando se compra um veículo, o primeiro pensamento de um indivíduo é tentar segurar o máximo possível em caso de sinistro. Para isso, para além da cobertura obrigatória de responsabilidade civil, as companhias oferecem a possibilidade de um seguro de danos próprios que cobre uma vasta possibilidade de coberturas extra. Não sendo esta uma cobertura obrigatória, o seguro de Danos Próprios é um seguro que pode cobrir três tipos de sinistros:

- Choque, Capotamento ou Colisão – Previne o cliente de qualquer acidente provocado por ele, resultante de um choque, capotamento ou colisão;
- Furto ou Roubo – Previne o cliente caso o seu carro sofra uma tentativa de assalto ou de roubo;
- Incêndio, Raio ou Explosão – Previne o cliente de qualquer sinistro que resulte num incêndio, raio ou explosão.

Qualquer uma destas garantias podem ser contratadas isoladamente, ou agrupadas em módulos, possibilitando a cobertura de um risco isolado ou conjunto de riscos.

Todavia, estes tipos de garantias também têm diversas exclusões, como:

- Se o prejuízo causado pelo sinistro for causado de forma intencional ou voluntária por parte do tomador do seguro, segurado, condutor, restantes ocupantes, ou qualquer pessoa que com eles coabite;
- Se os prejuízos ocorrerem quando o condutor do veículo seguro esteja com taxa de alcoolémia superior à legalmente exigida, ou efeito de estupefacientes;
- Se os prejuízos ocorrerem quando o condutor do veículo não esteja legalmente habilitado para conduzir;
- Se o sinistro ocorrer como consequência de fenómenos da natureza.

É de referir, também, que o valor que a companhia de seguros terá de pagar aquando de sinistros afetados por esta cobertura é o valor do veículo no início da anuidade. Este é obtido através da diferença entre o valor do veículo e a desvalorização do mesmo perante a tabela de desvalorização de Veículos da Companhia de Seguros.

4.4. Outros Danos

É também oferecido ao cliente a proposta de poder ter um seguro facultativo que cobre tipos de sinistros que não estão presentes nas coberturas anteriormente descritas. O seguro de Outros Danos é um seguro de coberturas facultativas que pode cobrir:

- Assistência em Viagem – Cobre qualquer problema que o segurado tenha durante uma viagem. É exemplo um furo de pneu ou avaria do motor;
- Acidentes Pessoais Ocupantes – Esta cobertura garante a segurança a todos os indivíduos presentes num veículo ao qual lhe foi atribuída culpa;

- Prejuízos Adicionais – Esta cobertura garante o pagamento por parte da seguradora de sinistros resultantes de danos materiais causados a bagagens de ocupantes em consequência direta de sinistros cobertos pela cobertura de Danos Próprios;
- Privação de Uso – Esta cobertura garante um valor de indemnização durante o período de reparação ou desaparecimento do veículo em consequência de um sinistro;
- Proteção Jurídica – Esta cobertura garante a defesa de um processo penal em consequência de um acidente de viação;
- Quebra Acidental de Vidros – Esta cobertura garante uma indemnização correspondente a gastos de substituição e montagem de para-brisas, oculo traseiro, vidros laterais ou qualquer tipo de vidro existente no automóvel, em caso de quebra ou rotura dos mesmos;
- Riscos Extraordinários – Esta cobertura garante os prejuízos diretamente provocados por deslizamentos ou aluimento de terras, derrocadas, abatimento de tuneis, pontes, queda de aeronaves, fenómenos sísmicos, ou qualquer acontecimento que não seja comum no dia a dia;
- Riscos Sociais e Políticos – Esta cobertura garante o pagamento por parte da seguradora de qualquer sinistro resultante de greves, tumultos, motins, terrorismo ou alterações da ordem pública.

4.5. Bónus-Malus

O sistema Bónus-Malus é uma forma das companhias seguradoras beneficiarem ou penalizarem os seus clientes no valor do prémio do seguro como consequência da prestação do segurado na anuidade anterior. Desta forma, Bónus é a redução percentual do prémio de algumas coberturas se o cliente não tiver nenhum sinistro da sua responsabilidade. Malus é o aumento percentual do prémio de algumas coberturas se o cliente tiver sinistros da sua responsabilidade na anuidade anterior.

No anexo A está representada a tabela Bónus-Malus em vigor e a sua forma de processamento. Numa determinada anuidade, o cliente está num determinado nível, e consoante a existência de sinistros pode mudar de nível. Inicialmente começa-se no nível O e a não ocorrência de sinistros faz com que mude para o nível imediatamente anterior. Quando ocorre uma mudança na tabela de Bónus-Malus, os clientes ficam com os valores da tabela que lhes foi fornecida no 1º contrato, não sofrendo alterações, ou seja, esta tabela é apenas para novos clientes.

Para efeito de bonificação ou agravamento do prémio a partir do sistema de Bónus-Malus, apenas se consideram sinistros que afetem alguma das seguintes coberturas:

- Danos Acidentais Sofridos pelo Veículo;
- Furto ou Roubo;
- Incêndio, Raio ou Explosão;
- Responsabilidade Civil Obrigatória ou Facultativa.

4.6. Fundo de Garantia Automóvel

Apesar de ser obrigatória a existência de um contrato de responsabilidade civil num veículo em Portugal, nem todos o possuem. Em consequência disto, criou-se um fundo de Garantia Automóvel com o objetivo de garantir o pagamento de indemnizações quer materiais quer corporais decorrentes de acidentes causados por veículos que não tenham seguro. Assim sendo, é possível acionar o fundo de garantia automóvel quando os sinistros causados tenham as seguintes características:

- Morte ou danos corporais onde o responsável seja conhecido e não tenha seguro válido;
- Quando o responsável de um sinistro seja desconhecido;
- Danos materiais quando o responsável pelo sinistro é conhecido e não tem seguro válido.

4.7. Regularização de Sinistros

De forma a facilitar a regularização de sinistros foi criado a chamada Indemnização Direta ao Segurado (IDS). A IDS é uma indemnização ao segurado por parte da sua seguradora aquando a ocorrência de um sinistro que não é da sua responsabilidade. Ou seja, é um acordo feito entre seguradoras, no âmbito do seguro automóvel, que permite ao tomador de seguro contactar a sua seguradora mesmo que não seja responsável pelo acidente. A sua seguradora adianta-se com o pagamento do sinistro e depois trata de tudo com a seguradora do responsável.

Este acordo já existe em praticamente todas as seguradoras em Portugal, com a aprovação da Associação Portuguesa de Seguradores.

Para que se possa regularizar um sinistro através do IDS, existem diversas condições que devem ser respeitadas:

- A Declaração Amigável de Acidente Automóvel tem de estar devidamente preenchida e assinada;
- O Sinistro apenas pode envolver dois veículos, sendo que a causa tem de ser colisão direta, não podendo estar alocado a danos corporais;
- A reparação dos danos materiais resultantes do acidente não seja superior a 15000€;
- O acidente tem de ocorrer no território Português.

5. Análise Estatística

5.1. Regressão

A regressão é uma das metodologias estatísticas mais interessantes de se estudar. Esta consiste na criação de modelos estatísticos que permitem estudar o comportamento de diversas variáveis. A partir da construção de um modelo estatístico podemos analisar o comportamento da variável resposta em função das variáveis explicativas. Existem vários tipos de regressão, contudo neste trabalho de projeto apenas vamos estudar a Regressão Logística.

5.1.1. Regressão Logística

Supondo que Y representa a variável independente ou variável resposta e X a variável dependente ou explicativa, na regressão linear admite-se que o valor médio condicional pode ser escrito como equação linear de x , ou seja, $E[Y|x] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k = \beta_0 + \sum_{i=1}^k \beta_i X_i$. A partir desta equação percebemos que $E[Y|x]$ pode tomar valores de $]-\infty; +\infty[$.

Um modelo de regressão logística é um modelo em que a variável dependente é binária, ou seja, pode tomar valores “0” ou “1”. Assim, se tomar o valor de 1 podemos dizer que a i -ésima observação possui a característica de interesse, ou seja, a característica que queremos estudar. Neste tipo de regressão, a variável dependente Y segue uma distribuição Bernoulli, com probabilidade de sucesso igual a $\pi(x_i)$, ou seja, é a probabilidade desconhecida da existência da característica em estudo, associada às características das covariáveis padronizadas.

X é uma matriz composta por constantes conhecidas, ou seja, não é uma matriz aleatória. Esta possui todas as variáveis independentes em estudo. Sendo assim cada observação y_i tem uma distribuição de acordo com a seguinte tabela:

Y_i	Probabilidade
1	$P(Y_i=1 x_i) = \pi(x_i)$
0	$P(Y_i=0 x_i) = 1-\pi(x_i)$

Pela definição de variáveis aleatórias discretas, sabemos que $E(Y_i|x_i) = \pi(x_i)$, ou seja, o valor esperado de Y_i representará a probabilidade de sucesso $\pi(x_i)$.

Um modelo de regressão logística tem a forma:

$$\pi(x_i) = P(Y_i = 1|X = x_i) = \frac{\exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik})}{1 + \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik})}$$

Através da função logit, podemos fazer uma alteração em $\pi(x)$ e obter uma nova equação.

Sendo $h(x) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$.

Utiliza-se esta função porque possui todas as propriedades de um modelo de regressão linear, ou seja:

- $h(x)$ pode ser contínua;
- $h(x)$ é linear nos seus parâmetros;
- $h(x)$ pode variar em todo o \mathbb{R} dependendo do domínio de x .

Dentro da regressão logística, há dois tipos de variáveis possíveis a entrar no modelo:

- Variáveis quantitativas: São aquelas que são numericamente mensuráveis, ou seja, que apresentam valores numéricos. É exemplo, a idade de um indivíduo em anos ou a antiguidade de carta de condução;
- Variáveis qualitativas: São variáveis que se baseiam em características das observações e que não podem ser calculadas numericamente. Este tipo de variáveis necessitam de ser tratadas de forma diferente às variáveis quantitativas, ou seja, para este tipo de variáveis tem de se fazer uma transformação e criar variáveis dummy. As variáveis dummy são variáveis categóricas que são transformadas em numéricas. Assim, a partir da criação das mesmas já é possível a construção de um modelo.

Teste de WALD

O teste de Wald é um teste que serve para ver se cada coeficiente da variável é significativamente diferente de zero. Após a estimação de cada parâmetro é possível obter-se uma resposta acerca da viabilidade da variável, ou seja, através da comparação entre a estatística de máxima verosimilhança do parâmetro $\hat{\beta}_j$ e a sua variância.

As hipóteses a testar são as seguintes:

$$H_0: \beta_j = 0 \text{ vs } H_1: \beta_j \neq 0, \quad j = 0, \dots, p, \text{ sendo } p \text{ as variáveis a testar}$$

A estatística de teste é dada por

$$W_j = \frac{(\hat{\beta}_j)^2}{\text{var}(\hat{\beta}_j)} \bigg|_{\text{sob } H_0} \chi^2_{(1)}$$

5.2. Indicadores de Comparação

5.2.1. Matriz de Confusão

Existem muitos métodos para se poder avaliar a qualidade de ajustamento de um modelo de regressão logística. Um deles é através da chamada Matriz de Confusão. Este ajuda a comparar os valores observados com os valores previstos no modelo, ou seja, re-classifica os valores da variável resposta de acordo com uma variável dicotômica, sendo esta variável construída à custa das probabilidades estimadas.

O modelo estima a probabilidade de cada observação tomar o valor “1”. Consequentemente a probabilidade ocorrência deste sucesso está no o intervalo $]0,1[$. Considerando o valor desta probabilidade, é necessário prever se a variável dependente toma o valor de “1” ou “0”. Temos então o chamado ponto de corte ou cut-off. Este é o valor para o qual se toma a decisão de qualidade do modelo. Se o valor estimado para cada Y_i for superior ao nosso ponto de cut-off então esta observação tomará o valor de “1”, se for inferior toma o valor de “0”.

Desta forma, têm-se para todas as observações os chamados valores previstos e os observados, permitindo assim avaliar a concordância entre os mesmos.

Pode-se então definir os seguintes conceitos:

- Verdadeiros Positivos (VP) – São as observações que o modelo previu que fossem positivas e que na verdade o são;
- Verdadeiros Negativos (VN) – São os valores previstos como negativos e que na realidade o são;
- Falsos Positivos (FP) – São os valores negativos que foram erradamente previstos como positivos, ou seja, o modelo previa que fossem positivos, mas observou-se que não o eram;
- Falsos Negativos (FN) – São os valores positivos que foram erradamente previstos como negativos, ou seja, são as observações que o modelo previa que fossem falsas, mas na realidade são positivas.

Uma vez classificados todos os pares de valores observados e previstos, pode se apresentar os resultados na forma de uma matriz, a Matriz de Confusão:

		Valor Verdadeiro (confirmado por análise)	
		positivos	negativos
Valor Previsto (predito pelo teste)	positivos	VP Verdadeiro Positivo	FP Falso Positivo
	negativos	FN Falso Negativo	VN Verdadeiro Negativo

Figura 5.1 - Matriz de Confusão ¹

¹ Disponível em: <http://developerdeveloper.blogspot.pt/2013/11/matriz-confusao.html>; Acesso em Abril 2017.

Um modelo torna-se melhor à medida que o número de falsos positivos (FP) e o número de falsos negativos (FN) forem menores. No entanto, somos nós que os definimos através do ponto de corte explicado anteriormente. Sendo assim pode-se alterar e definir diferentes pontos de corte.

Uma vez construída a matriz de confusão, define-se as seguintes probabilidades:

Sensibilidade (**S**): proporção de verdadeiros positivos na amostra, ou seja, no caso que estamos a estudar, a proporção de clientes que renovaram com a companhia e que o modelo previa que renovassem, sobre o total de indivíduos que renovaram com a companhia. É expressa da seguinte forma:

$$S = \frac{VP}{VP + FN}$$

Especificidade (**E**): proporção de verdadeiros negativos, ou seja, a proporção de clientes que o modelo previa que não renovassem e que realmente não renovaram, sobre o total de clientes que não renovaram. É expressa da seguinte forma:

$$E = \frac{VN}{VN + FP}$$

(1-Sensibilidade): proporção de falsos negativos:

$$1 - S = \frac{FN}{VP + FN}$$

(1-Especificidade): proporção de falsos positivos:

$$1 - E = \frac{FP}{FP + VN}$$

Um bom modelo é aquele que tanto a proporção de falsos positivos ($1-E$) como a proporção de falsos negativos ($1-S$) são baixas. Podemos também dizer que a percentagem de observações corretamente classificadas é dada por: $\frac{VP+VN}{n}$, sendo n o número total de observações.

5.2.2. Curva ROC e Área Abaixo da Curva

Outro critério de escolha de modelos é obtido através da Curva ROC (Receiving Operating Characteristic Curve). Esta é uma forma de expressar-se graficamente a sensibilidade e o inverso da especificidade para um conjunto de pontos de corte (Cut Points) possíveis. Sendo assim, através da matriz de confusão, calcula-se a sensibilidade e especificidade para os diferentes pontos de corte. A tabela seguinte apenas servirá de exemplo para a construção gráfica:

Tabela 5.1 – Diferentes pontos de corte

Ponto de Corte	Sensibilidade	Especificidade	1 - Especificidade
0,1	95,4%	12,2%	87,8%
0,15	91,3%	36,6%	63,4%
0,2	86,6%	54,7%	45,3%
0,25	81,0%	68,2%	31,8%
0,3	69,0%	74,9%	25,1%
0,35	55,1%	83,1%	16,9%
0,4	42,1%	88,7%	11,3%
0,5	23,0%	94,7%	5,3%
0,6	10,5%	99,1%	0,9%
0,7	2,5%	100,0%	0%

Os valores obtidos na tabela anterior encontram-se representados no seguinte gráfico.

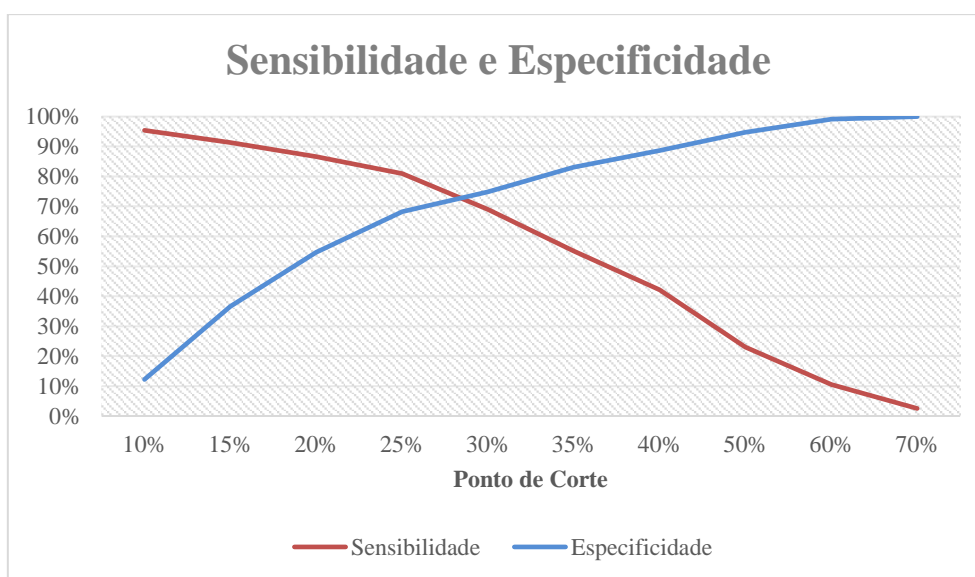


Gráfico 5.1 - Sensibilidade e Especificidade

Pela análise do gráfico anterior podemos verificar que o ponto que maximiza a reta da sensibilidade e da especificidade é o ponto de interseção de ambas, ou seja, no ponto de corte de 0,28.

Através da Tabela 5.1 podemos construir uma curva ROC. A figura seguinte apresenta a mesma:

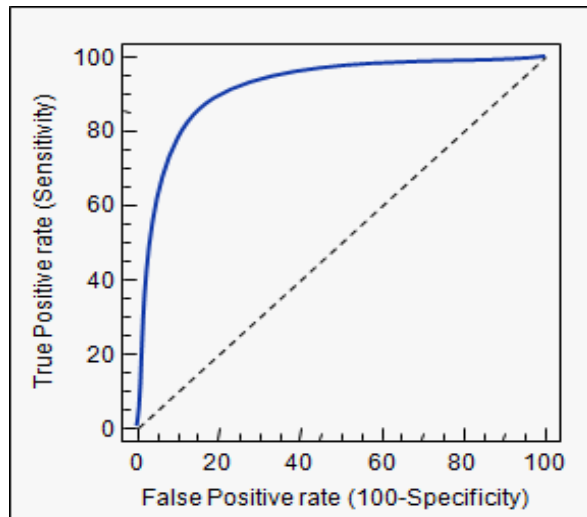


Figura 5.2 - Curva ROC²

Note-se que esta figura é apenas ilustrativa e não correspondeu à realidade.

Esta figura permite-nos visualizar a performance de uma variável dependente binária. Mas o verdadeiro interesse na curva ROC é o cálculo da área que está abaixo dela. Seja AUC (Area Under the Curve) a área abaixo da curva ROC, é possível classificar o modelo em 5 níveis de discriminação:

- [0.5-0.6] : Não há discriminação;
- [0.6-0.7] : Discriminação Fraca;
- [0.7-0.8] : Discriminação Aceitável;
- [0.8-0.9] : Discriminação Excelente;
- [0.9-1] : Discriminação Excecional.

Assim, quanto maior o valor da área abaixo da curva ROC melhor o modelo explica os dados.

Desta forma aquando a comparação de modelos, o modelo que apresente maior valor AUC melhor explica os dados e consequentemente é mais significativo.

² Disponível em : https://www.medcalc.org/manual/_help/images/roc_intro3.png;. Acesso em Abril 2017

5.2.3. Critério de Informação de Akaike e Bayesiano

Outro critério possível de comparação de modelos é o chamado Critério de Informação de Akaike (AIC). O AIC foi criado em 1973 e através da informação de Kullback-Leibler baseada na função de máxima verosimilhança consegue comparar dois modelos. O critério de informação de Akaike é calculado como:

$$AIC = -2 \log(L_p) + 2p$$

Em que:

L_p é a função de máxima verosimilhança do modelo;

p é o número de variáveis explicativas consideradas no modelo;

Este critério admite que haja um modelo “real” desconhecido que descreve a 100% os dados, pelo que com a ajuda deste tipo de critério podemos comparar de entre os diversos modelos, o que melhor justifica os dados, ou seja, o que minimiza a divergência de Kullback-Leibler. Quanto menor o valor de AIC melhor o modelo está ajustado.

Para corrigir modelos que tenham poucas observações na amostra, pode-se utilizar o critério de informação de Akaike corrigido (AICc). Este é calculado como $AIC + \frac{2p(p+1)}{n-p-1}$

Sendo p o número de variáveis explicativas do modelo e n o número de observações da amostra.

Também existe o critério de informação Bayesiano (BIC).

Este é calculado como $BIC = -2 \log(L_p) + 2p * \log(n)$

Em que:

L_p é a função de máxima verosimilhança do modelo

p é o número de variáveis explicativas do modelo

n é o número de observações da amostra.

Estes métodos apresentados anteriormente são métodos que têm uma característica em comum: penalizam os modelos que possuem maior numero de variáveis independentes.

5.3. Multicolinearidade

A multicolinearidade entre variáveis é um problema muito comum nas regressões que consiste na existência de variáveis independentes que possam estar a ser fortemente explicadas por outras. Como consequência, a existência de multicolinearidade resulta numa apresentação de um erro-padrão bastante elevado, ou a impossibilidade de estimação de um modelo.

De entre muitos métodos possíveis para o tratamento deste problema, optou-se pelo método de análise dos fatores de inflação da variância (VIF) pois foi um método estudado ao longo de unidades curriculares. Esta mede o quanto a variância do coeficiente $\hat{\beta}_j$ é inflacionado pela sua colinearidade.

O VIF da j -ésima variável é calculado da forma $VIF_j = \frac{1}{1-R_j^2}$ $j = 1 \dots p$. Para obter este valor calcula-se a matriz R^{-1} , isto é, a inversa da matriz de correlações das variáveis independentes, e este é o j -ésimo elemento da diagonal principal desta matriz.

Verificado o VIF das diversas variáveis independentes, podemos chegar a três conclusões:

- Se $VIF=1$ então a variável não está correlacionada;
- Se $1 < VIF < 10$ então a variável está moderadamente correlacionada;
- Se $VIF \geq 10$ então a variável está altamente correlacionada.

Se existirem variáveis independentes com valor de VIF maiores ou iguais a 10, então ter-se-á de as retirar da análise de regressão. Sendo assim, perante a exclusão da variável que possuir maior valor de VIF, repete-se o mesmo processo até todas as variáveis independentes não estarem altamente correlacionadas.

5.4. Tipos de Seleção Possíveis

No mercado segurador e para efeitos de retenção, grande parte dos modelos que as empresas utilizam são derivados de modelos de regressão, de forma a tentar perceber o comportamento dos seus clientes. Contudo existe variáveis que não podem estar presentes nos modelos, tendo de ser excluídas pois do ponto de vista comercial não podem ser consideradas num modelo. Por outro lado, do ponto de vista estatístico e tendo em conta a amostra disponível, essas variáveis devem ser incluídas, pois são variáveis importantes a explicar o risco.

Adicionalmente, o analista deve evitar incluir no seu modelo variáveis correlacionadas entre si que podem levar a uma distorção dos resultados finais. Para que isto não aconteça, antes de proceder ao desenvolvimento do modelo, é necessário efetuar análises preliminares a variáveis de modo a excluir variáveis independentes que possam condicionar o modelo, provocando problemas de dispersão ou de multicolinearidade.

Quando se quer construir um modelo, tem-se por base um método de seleção de variáveis. O modelo criado ao longo deste trabalho foi contruído com base no método de seleção regressiva. Inicialmente fez-se uma análise entre este método e o método de seleção progressivas, e verificou-se que este apresentava melhores resultados.

5.4.1. Método de Seleção Regressiva

O Método de Seleção Regressiva, ou Backward Elimination, tal como o seu nome indica, é um método de exclusão de variáveis que visa passo a passo a eliminação delas, ou seja, o modelo começa com todas as variáveis independentes e em cada iteração vai-se excluindo a variável menos significativa, ou seja, a variável independente que apresente menor relação com a variável dependente. Quando esta variável é identificada, será excluída e repete-se o teste sem a mesma, até que todas as variáveis sejam significativas.

Desta forma, o processo pode ser resumido nos seguintes passos:

Passo 1: Começar com todas as (n) variáveis disponíveis;

Passo 2: Calcular o p-value através da estatística F das variáveis selecionadas;

Passo 3: Identificar a variável com maior p-value;

Passo 4: Decidir quando à exclusão da variável:

Passo 4.1: Se o maior p-value for superior ou igual ao nível de significância escolhido, deve retirar-se essa variável e repetir-se o processo desde o passo 2;

Passo 4.2: Se o maior p-value obtido for inferior ao nível de significância escolhido significa que todas as variáveis testadas são significativas e o processo termina.

No fim do método de seleção regressiva, fica-se com $n - p$ variáveis no modelo, sendo que p é o número de variáveis excluídas.

6. Construção do Modelo

Quando queremos construir um modelo, e se considera um conjunto alargado de variáveis, deve-se realizar numa primeira fase, a análise univariada a todas as possíveis variáveis a introduzir no modelo. Assim, iniciamos este capítulo com os todos os critérios de seleção das apólices, passando seguidamente à análise descritiva das variáveis em estudo. Posteriormente detalha-se a criação do modelo.

Devido ao elevado número de variáveis e observações (64.082), utilizou-se um Software bastante desenvolvido que tem um papel fundamental neste capítulo. Tanto a fase de análise univariada como a criação do modelo de regressão logística, foram realizadas em SAS Enterprise Guide. A opção deste software residiu no facto de ser atualmente uma ferramenta bastante completa e poderosa.

6.1. Escolha de Observações

Com o objetivo de estimar a probabilidade de retenção, apenas se consideraram as apólices que tanto à data de renovação como no ano anterior à mesma, possuíssem o mesmo risco. Assim, consideraram-se apenas:

- Apólices que a companhia propôs renovação de contrato;
- Apólices com proposta de renovação entre Janeiro e Agosto de 2016;
- Apólices que não mudaram de agente;
- Incluiu-se apólices individuais e de empresas;
- Apólices que não pagassem o prémio em forma de prestação (mais de 4 vezes num ano);
- Apólices em que não houve mudança de veículo, nem alteração nas coberturas ou franquias (muitas vezes acontece na cobertura de danos próprios);
- Apólices em que não houve mudança no capital seguro para a cobertura Responsabilidade Civil (mudanças de 6 milhões para 50 milhões, ou vice-versa).

Note-se que se retiraram:

- Apólices com sinistros da nossa responsabilidade, porque o prémio da anuidade seguinte ia ser superior ao da anuidade anterior;
- Apólices de empregados da Companhia;
- Apólices temporárias.

Estas exclusões/seleções são determinantes para termos um universo de dados onde no final o conjunto de resultados e conclusões não sejam enviesadas e demonstrem claramente o comportamento do Cliente no momento da renovação. Um objetivo prioritário é considerarmos apenas apólices que tenham exatamente o mesmo perfil de risco entre a ultima anuidade e a renovação seguinte. Com estes critérios de seleção garante-se a análise comportamental do cliente face ao novo prémio proposto, uma vez que são excluídas todas as apólices que serão agravadas

por razões externas ao processo normal de renovação do risco, como é o caso de agravamentos por sinistralidade ou alteração do preço por alteração do veículo.

6.2. Preparação dos Dados

Numa primeira fase, e tendo a base de dados global fornecida pela Generali, fez-se uma filtragem dos dados, de forma a não haver informações incoerentes. Observou-se todas as variáveis independentes ao ponto de não aparecerem observações discordantes. É exemplo a idade do segurado. Neste caso eliminou-se observações ao qual os segurados tivessem menos de 18 anos, ou mais de 100 anos (visto que provavelmente seriam erros). Excluiu-se também as observações que não tinham idade do segurado ou idade de carta de condução.

A base de dados fornecida contém as seguintes 24 variáveis independentes:

- Zona geográfica;
- Tipo de cliente;
- Grupo de agentes;
- Mês de renovação;
- Tipo de veículo;
- Existência da cobertura de quebra isolada de vidros;
- Existência da cobertura de danos próprios;
- Franquia da cobertura de danos Próprios;
- Bónus-Malus da Apólice;
- Prémio pago na anuidade de 2015;
- Prémio proposto para a anuidade de 2016;
- Variação entre o prémio proposto e o prémio pago em 2015 (em Percentagem);
- Frequência de pagamento;
- Marca do veículo;
- Tipo de combustível;
- Rácio entre peso e potência;
- Distrito;
- Valor do veículo;
- Idade do condutor;
- Idade do segurado;
- Idade do veículo;
- Antiguidade da carta de condução;
- Antiguidade da apólice;
- Retenção média do agente.

6.3. Análise Univariada

6.3.1. Transformação de dados

Para tentar evitar a divulgação de informações confidenciais, os dados da companhia foram sujeitos a uma transformação, pelo que, segue em seguida um pequeno exemplo da transformação:

Tabela 6.1 - Taxa de Retenção Por Tipo de Veículo

Tipo de Veículo	Taxa de Retenção	Taxa de Retenção Base 100
Ligeiro	79%	96%
Misto	79%	95%
Caminheta	75%	91%
TT - Ligeiro	79%	96%
TT - Misto	80%	97%
TT - Caminheta	82%	100%
Comercial	73%	89%
Auto Caravanas	79%	96%

Nota: Os valores apresentados na tabela anterior não são na sua totalidade valores reais da companhia.

Como o veículo com maior Taxa de Retenção é a TT – Caminheta, então será este o veículo de referência. Para se obter uma taxa de retenção Base 100, vai-se fazer um quociente entre o valor da taxa de retenção e o valor da taxa de retenção de maior valor, neste caso 82%. Para os veículos ligeiros, como têm uma taxa de retenção de 79%, então $\frac{79\%}{82\%} = 96\%$. Todos os valores foram arredondados às unidades.

Partimos assim do seguinte gráfico:

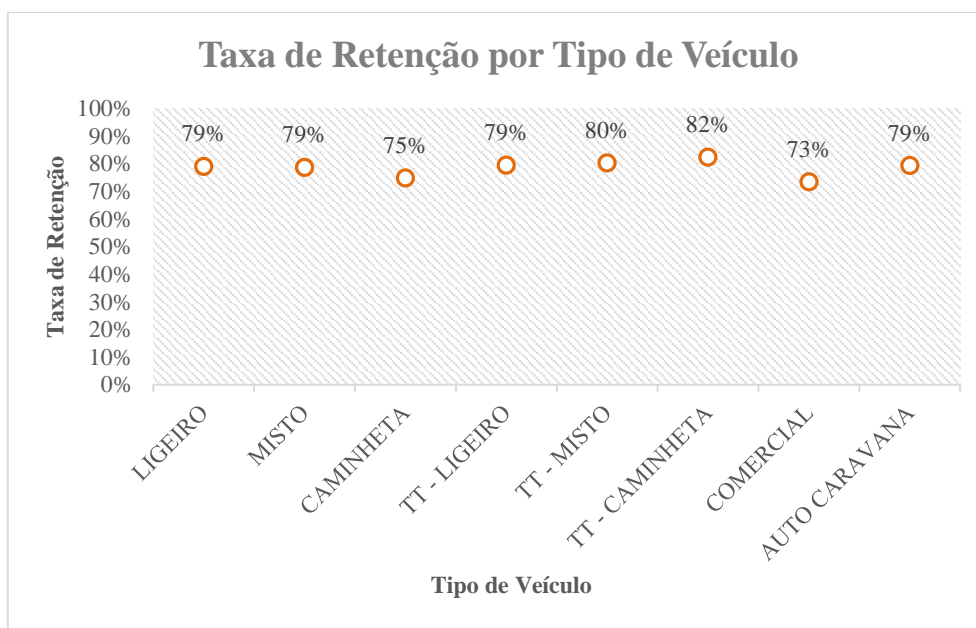


Gráfico 6.1 - Taxa de Retenção por Tipo de Veículo

Fazendo a transformação dos dados obtêm-se o seguinte gráfico:

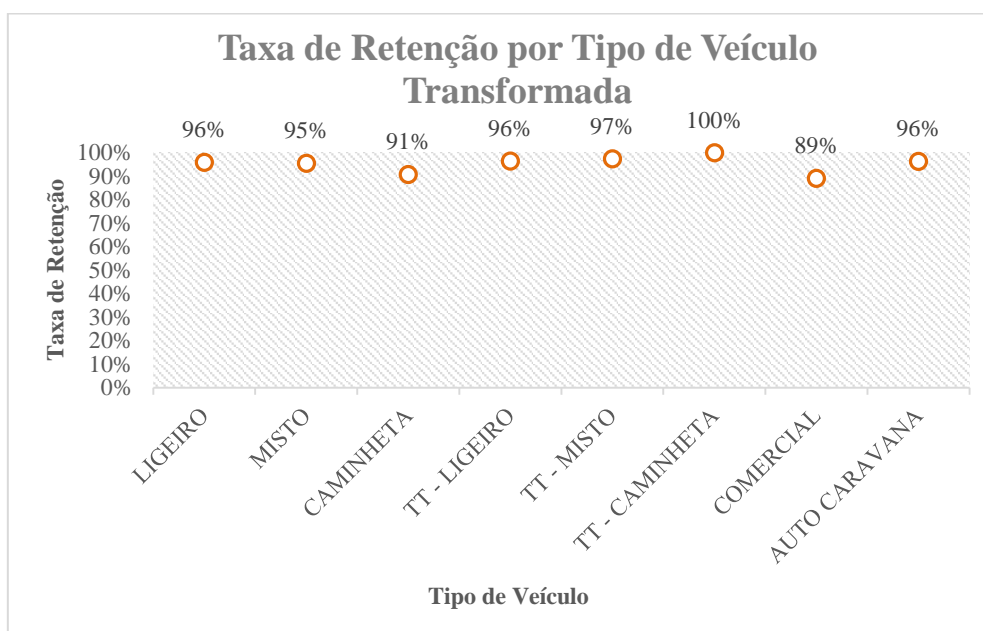


Gráfico 6.2 - Taxa de Retenção por Tipo de Veículo Base 100

Embora se tenha feito uma transformação nos dados (*Gráfico 6.2*), e se tenham obtidos outros valores de retenção, consegue-se visualizar a mesma tendência e chegar às mesmas conclusões que no *Gráfico 6.1*.

6.3.2. Análise Descritiva

Criou-se um pequeno programa no software utilizado (*Sas Enterprise Guide*) que nos permitiu criar as novas taxas de retenção. O programa segue no anexo B. Todos os gráficos possuem a exposição, valores representados em colunas, cujo valor de exposição se encontra na parte direita dos gráficos. Os gráficos para as variáveis em estudo são:

- Zona Geográfica

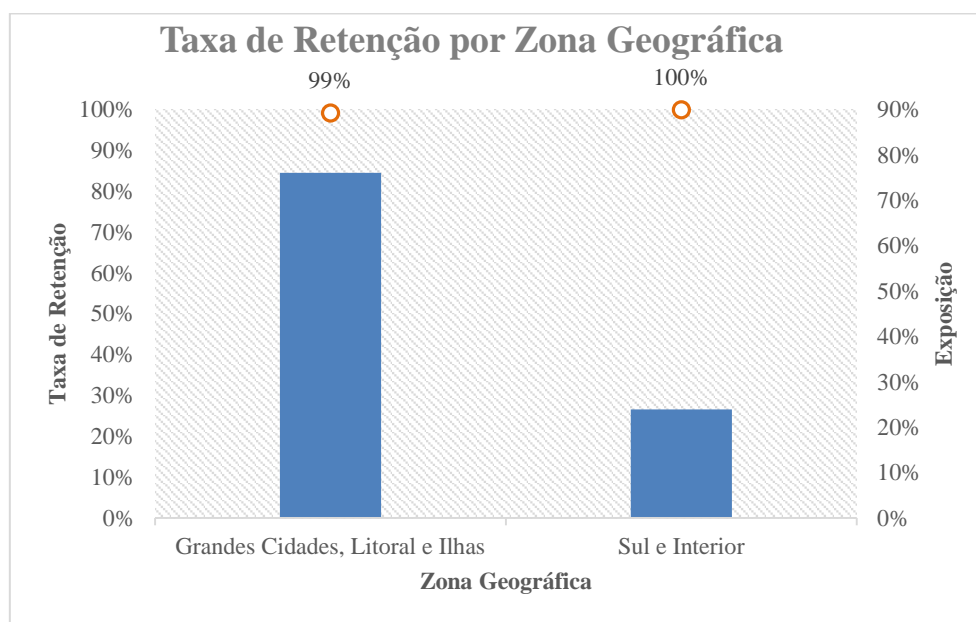


Gráfico 6.3 - Taxa de Retenção por Zona Geográfica

O Gráfico 6.3 mostra a variação da taxa de retenção por zona Geográfica. Esta variável divide Portugal em dois grupos (Grandes cidades, Litoral e Ilhas; Sul e Interior). O gráfico mostra-nos que a taxa de retenção é maior no sul e interior que nas grandes cidades, litoral e ilhas, embora se trate de uma pequena diferença. Um fator que pode explicar esta variação é a proximidade entre o cliente e o seu agente.

- Tipo de Cliente

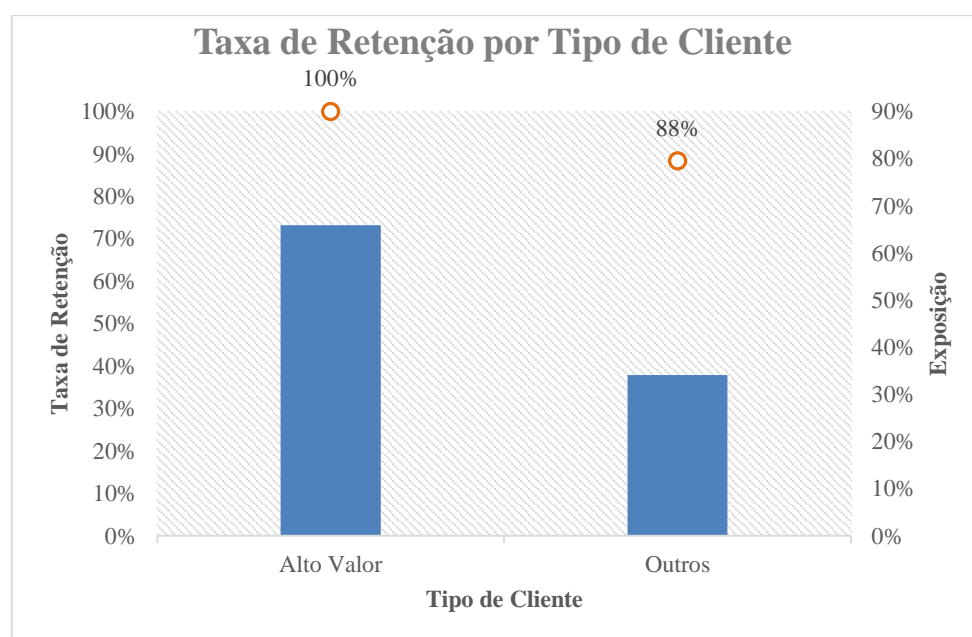


Gráfico 6.4 - Taxa de Retenção por Tipo de Cliente

O Gráfico 6.4 mostra como varia a retenção por tipo de cliente. A companhia divide os clientes em dois tipos: os de alto valor são aqueles que têm diversas características que a empresa considera importantes para o negócio e os que não têm estas características serão, por exclusão de partes, do outro grupo. Este gráfico mostra-nos que a taxa de retenção é maior para os clientes de alto valor.

- Grupo de Agentes

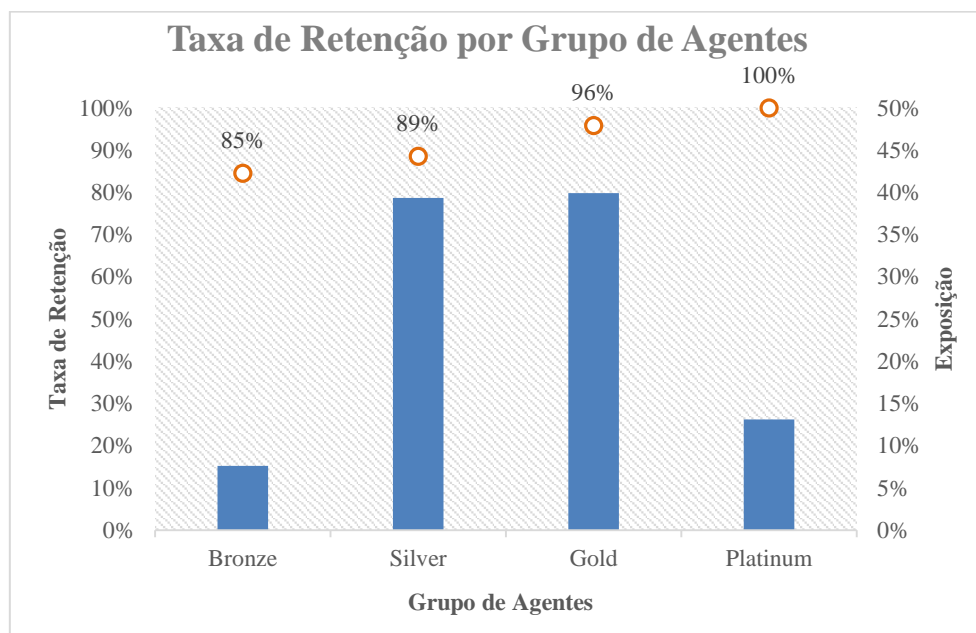


Gráfico 6.5 - Taxa de Retenção por Grupo de Agentes

O Gráfico 6.5 demonstra como varia a retenção dependente do tipo de agente. A companhia divide os agentes em 4 grupos de acordo com o seu volume de prémios e resultado que geram. Os agentes estão classificados como Bronze, Silver, Gold e Platinum, sendo que os Bronze representam o nível mais baixo e os Platinum o melhor. Sendo assim, de acordo com o gráfico anterior verifica-se que os agentes que estão no grupo dos Platinum apresentam uma maior taxa de retenção que os restantes. É normal que à medida que o nível de agente aumente, a sua taxa de retenção também aumente.

- Mês de Renovação

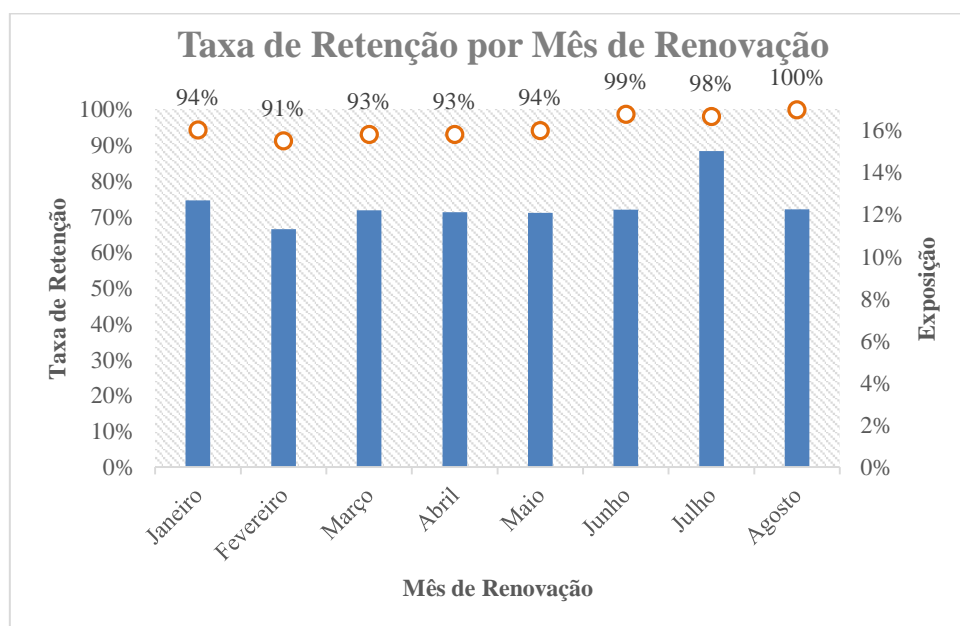


Gráfico 6.6 - Taxa de Retenção por Mês de Renovação

O Gráfico 6.6 demonstra como varia a taxa de retenção de acordo com o Mês de Renovação da apólice. É de notar que apenas foram consideradas as apólices que tiveram proposta de renovação nos primeiros 8 meses do ano. Podemos verificar que nos meses de maior calor (Junho, Julho e Agosto) a taxa de renovação é superior aos meses mais frios (Janeiro, Fevereiro e Março).

- Tipo de Veículo

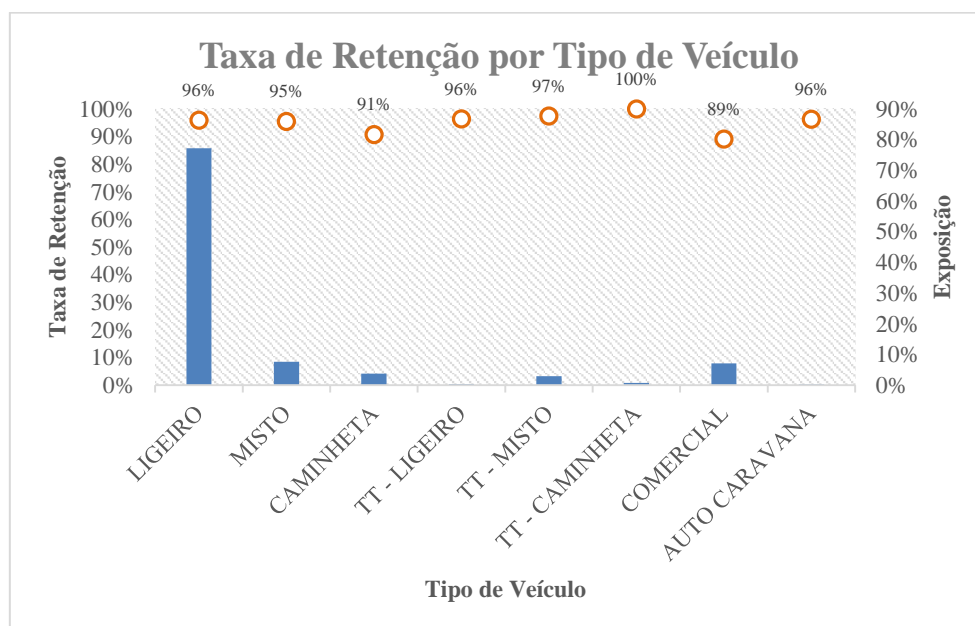


Gráfico 6.7 - Taxa de Retenção por Tipo de Veículo

O Gráfico 6.7 mostra-nos a variação da taxa de retenção por Tipo de Veículo. Podemos verificar que os veículos TT – Caminheta apresentam uma taxa de renovação superior aos outros tipos de veículos. Podemos também verificar que não existe uma grande diferença entre os Veículos Ligeiros, Mistos TT-Ligeiro, TT-Misto e Auto Caravanas. É de referir que TT significa todo o terreno.

- Cobertura de quebra isolada de vidros

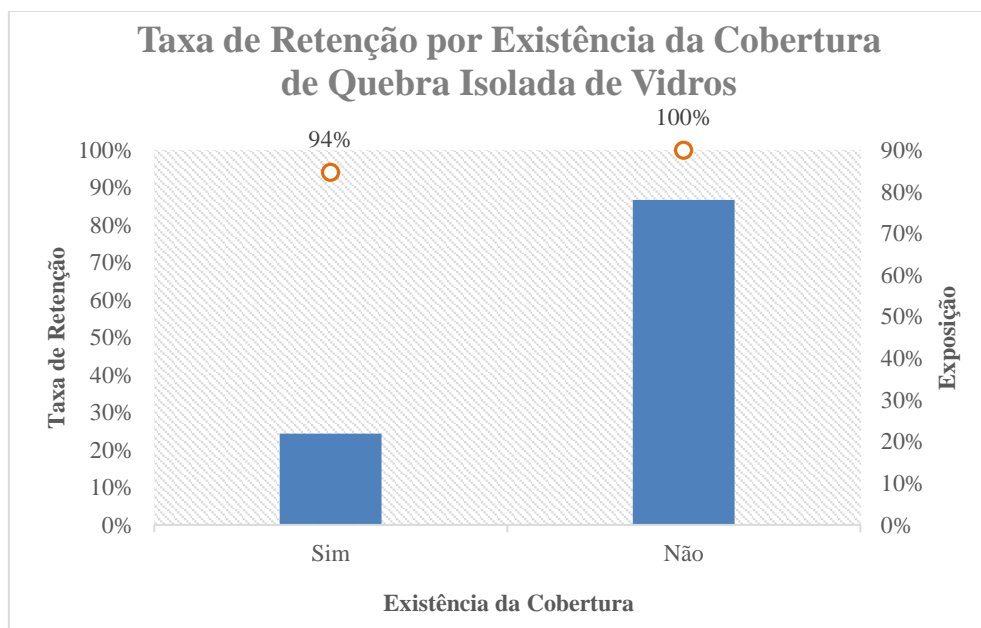


Gráfico 6.8 - Taxa de Retenção por Existência da Cobertura de Quebra Isolada de Vidros

O Gráfico 6.8 demonstra a variação da taxa de retenção pela existência de franquias da cobertura de quebra de vidros. Caso um cliente tenha esta cobertura terá uma franquia de 50 €. Com base neste gráfico, podemos afirmar que os clientes que não têm esta cobertura apresentam uma maior taxa de retenção que os clientes que a possuem, embora a diferença não seja muito acentuada. Pode-se dever ao facto de não ser uma cobertura gratuita

- Existência da Cobertura de Danos Próprios

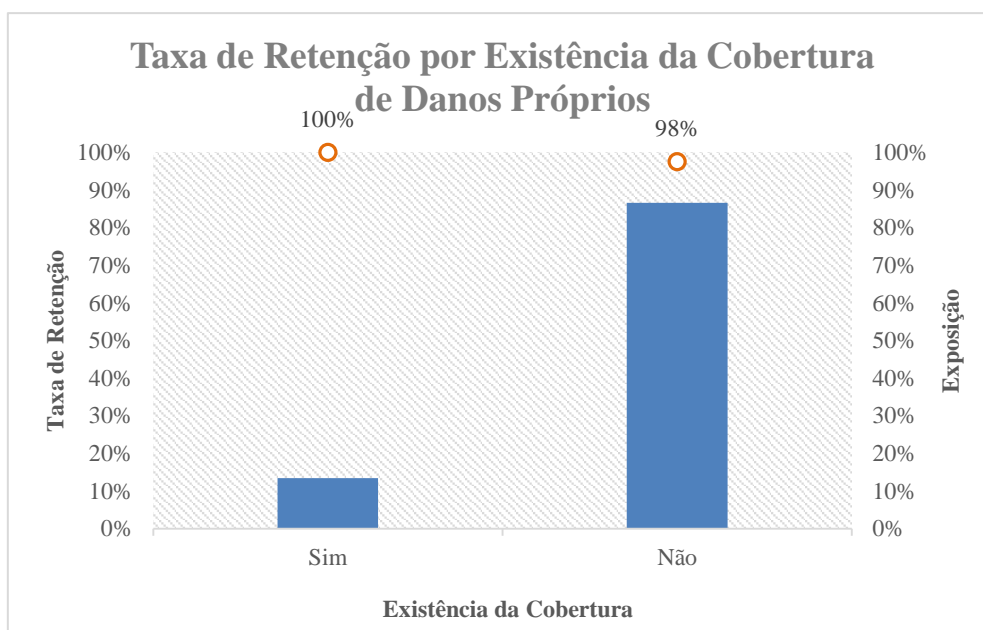


Gráfico 6.9 - Taxa de Retenção por existência da cobertura de Danos Próprios

O Gráfico 6.9 mostra a variação da taxa de retenção de acordo com a existência da cobertura de Danos Próprios na apólice. Quem tem esta cobertura, apresenta uma taxa de retenção superior face aos restantes. Isso indica-nos que se um cliente tiver a cobertura de danos próprios têm uma maior probabilidade de renovação.

- Franquia da Cobertura de Danos Próprios

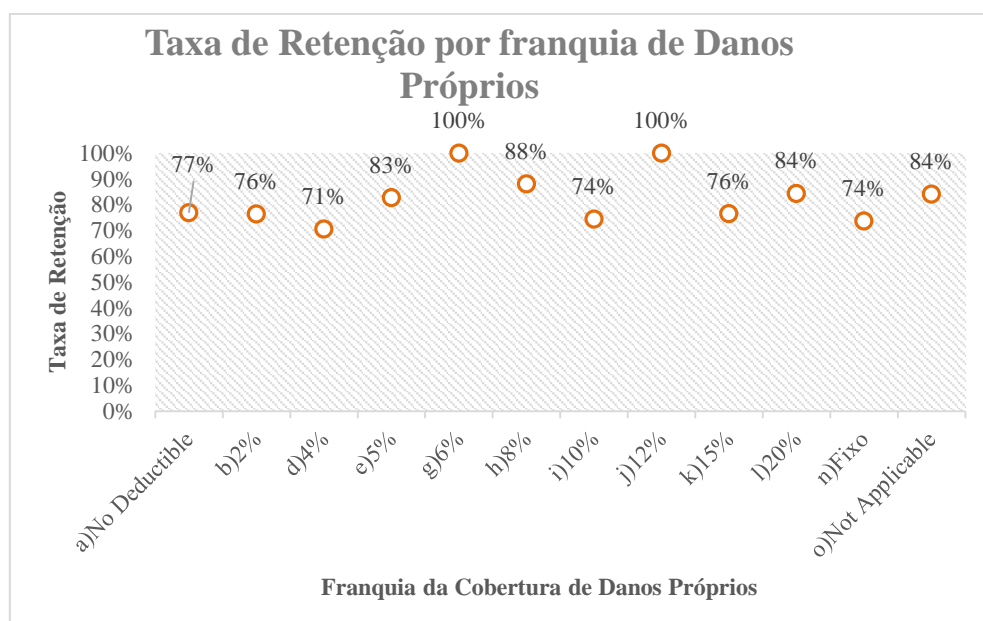


Gráfico 6.10 - Taxa de Retenção por franquia de Danos Próprios

O Gráfico 6.10 mostra a variação da taxa de retenção pela franquia da cobertura de Danos Próprios. E verifica-se que não segue um padrão linear, ou seja, com o aumento da franquia a pagar tanto tem-se subidas como diminuições na probabilidade de renovação.

- Bónus – Malus

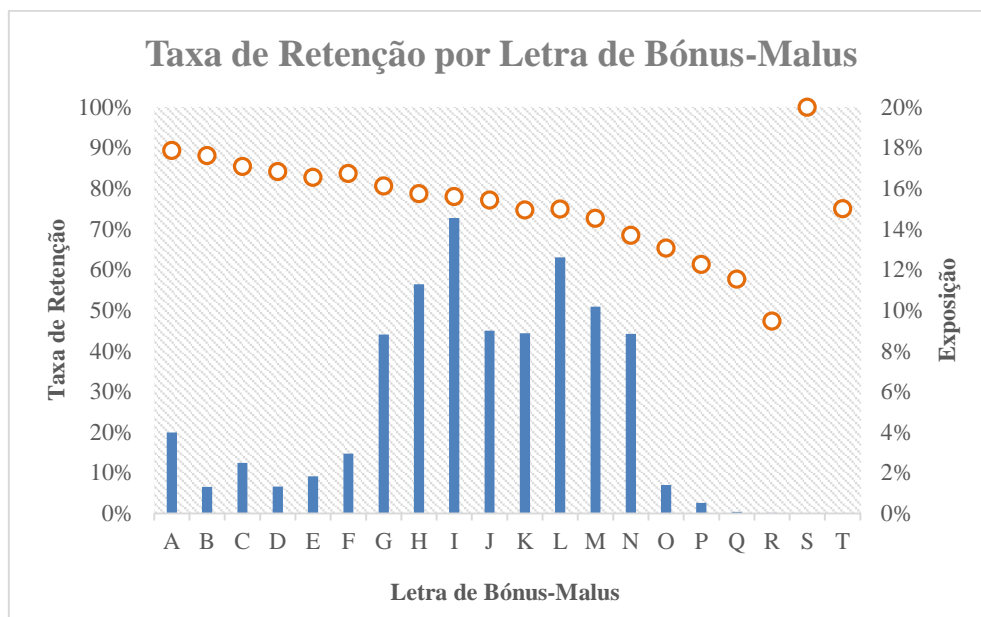


Gráfico 6.11 - Taxa de Retenção por Letra de Bónus-Malus

O Gráfico 6.11 mostra-nos a variação da taxa de retenção de acordo com a letra Bónus-Malus da apólice e indica-nos que as apólices com melhor letra Bónus-malus (A) apresentam uma taxa de retenção superior. É de reparar que à medida que a letra de Bónus-Malus diminui, a taxa de retenção segue a mesma tendência até à letra R. As letras S e T são caracterizadas por terem muito poucas apólices, o que causa algum ruído.

- Prémios

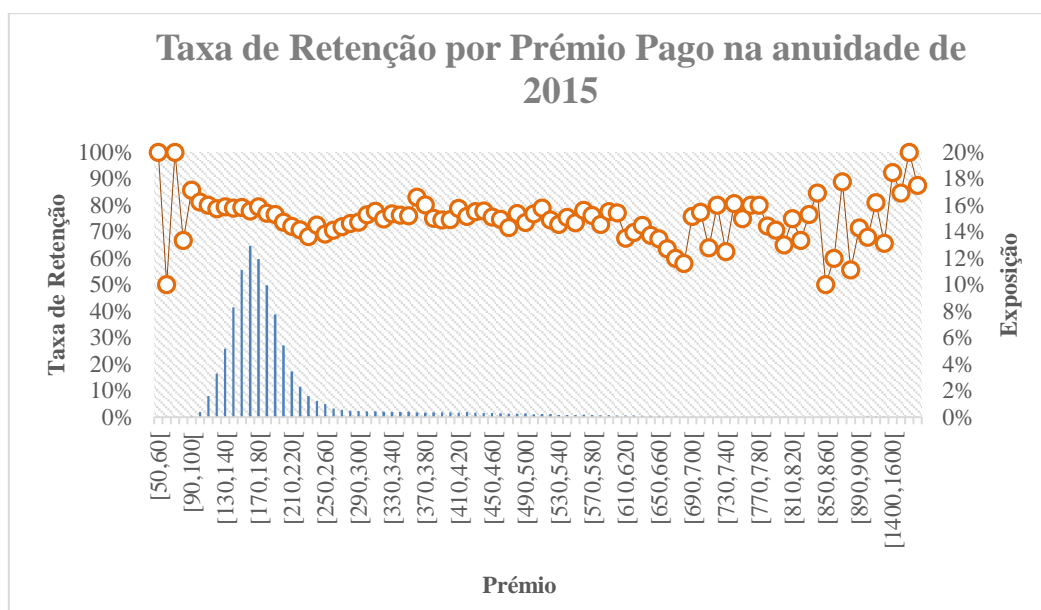


Gráfico 6.12 - Taxa de Retenção por prémio pago na anuidade de 2015

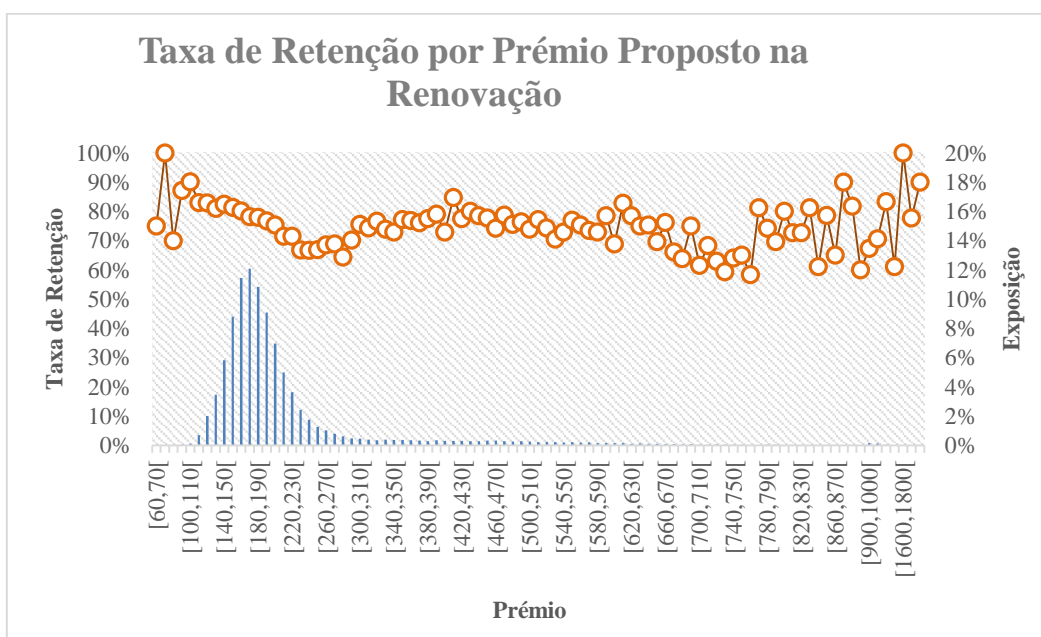


Gráfico 6.13 - Taxa de Retenção por Prémio proposto

Em relação ao *Gráfico 6.12*, este mostra-nos como varia a taxa de retenção de acordo com o valor do Prémio pago na anuidade de 2015. É de verificar que à medida que o prémio é superior, menor é a probabilidade de renovação do cliente. O *Gráfico 6.13* demonstra como varia a taxa de retenção de acordo com o prémio proposto na renovação. Este gráfico segue a mesma tendência do gráfico anterior. Conclui-se então que as pessoas são muito sensíveis ao preço do seu contrato de seguro, renovando mais facilmente se propuserem um valor menor de prémio.

- Variação do Prémio em Percentagem

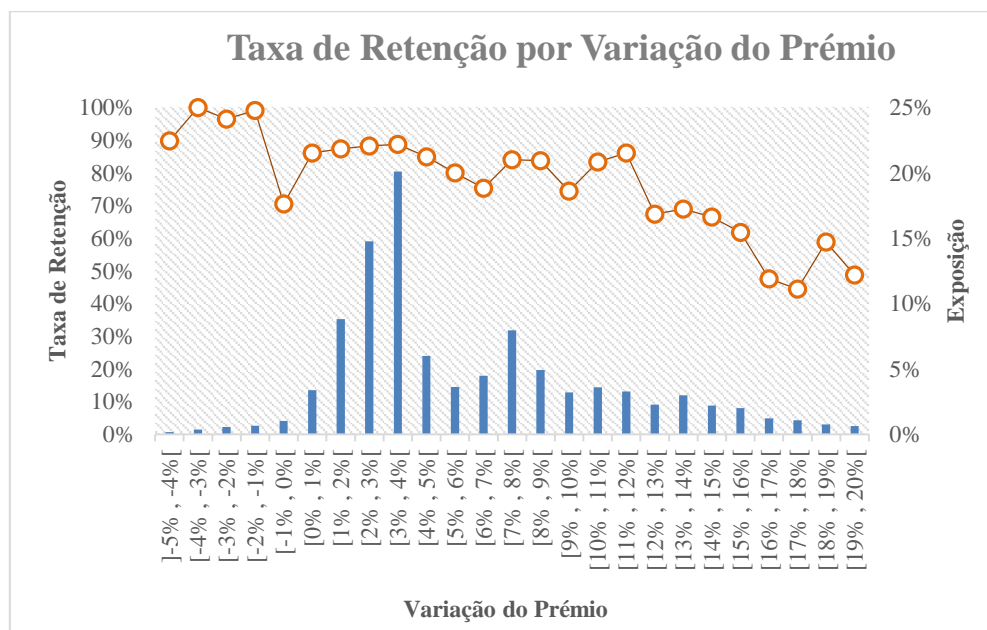


Gráfico 6.14 - Taxa de Retenção por Variação do Prémio em Percentagem

O Gráfico 6.14 mostra-nos como varia a taxa de retenção de acordo com a variação do Prémio em Percentagem.

Muitas vezes ao olharmos para o aumento/diminuição de prémios, este não nos indica se este aumento/diminuição é de facto grande/pequeno em termos relativos. Sendo assim, este gráfico mostra-nos a variação do prémio proposto para a renovação e o prémio pago em 2015, em percentagem. Verifica-se que quando há uma diminuição do prémio, a taxa de retenção tem tendência a ser maior que quando há um aumento. Mas dentro dos aumentos de prémios também é possível observar-se que quem sofre um maior aumento tem uma menor tendência a renovar.

- Frequência de Pagamento

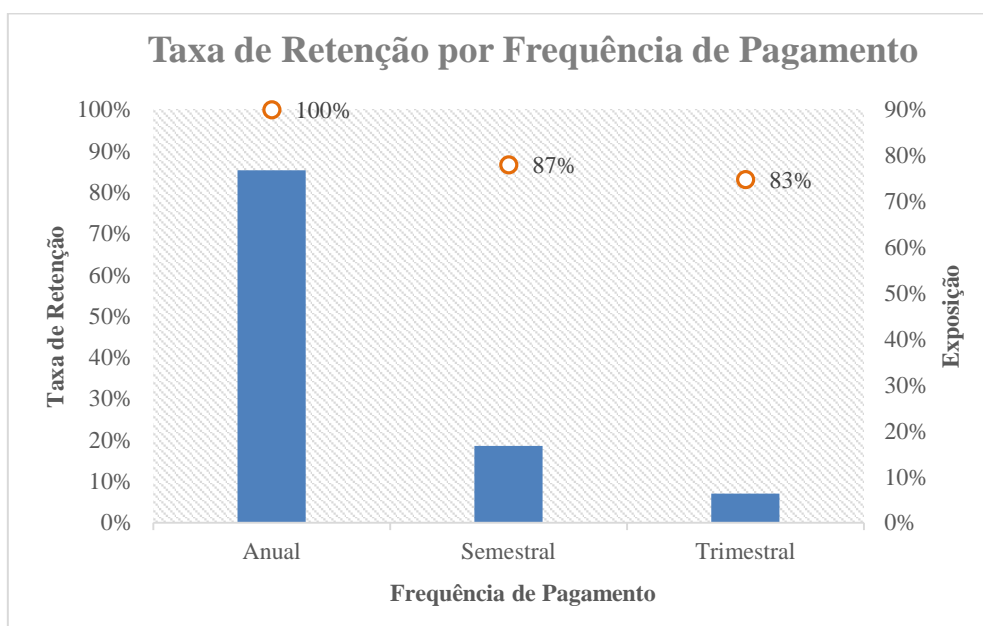


Gráfico 6.15 - Taxa de Retenção por Frequência de Pagamento

O *Gráfico 6.15* representa a variação da taxa de retenção de acordo com a frequência de pagamento. Nesta análise apenas foram apenas consideradas apólices que tinham três tipos de pagamento: Os que pagavam tudo de uma vez (anual), os que pagavam duas vezes num ano (semestral), ou os que pagavam no máximo quatro vezes (trimestral).

Como é de esperar, a taxa de retenção é maior para apólices que pagam tudo de uma vez. É muito comum no mercado segurador que quando o pagamento é feito por diversas vezes, os clientes tenham tendência para que ao mínimo conflito com a entidade seguradora, cancelar as suas apólices. É de verificar que quem paga 4 vezes num ano tem uma probabilidade de renovação inferior a quem o faz semestralmente.

- Marca

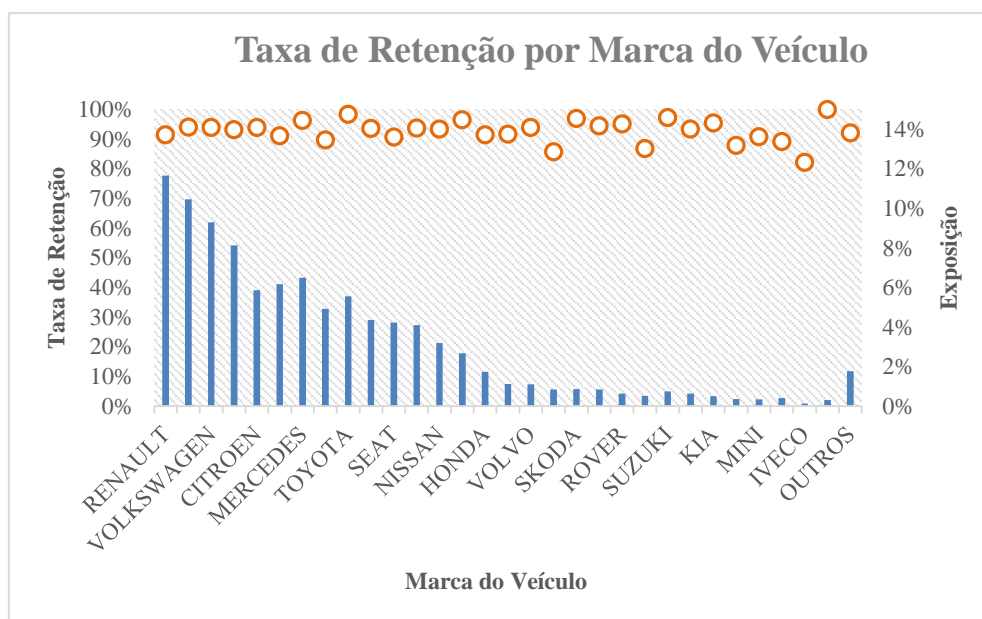


Gráfico 6.16 - Taxa de Retenção por Marca do Veículo

O Gráfico 6.16 é referente à variável Marca do Veículo. Ao observar-se o mesmo, verifica-se que há oscilações entre as marcas dos veículos enquanto que os carros de Gama Alta apresentam uma maior taxa de retenção em relação aos outros. As marcas mais comuns no Mercado português (Renault, Opel, Ford, Peugeot, Volkswagen) apresentam taxas equivalentes.

- Tipo de Combustível

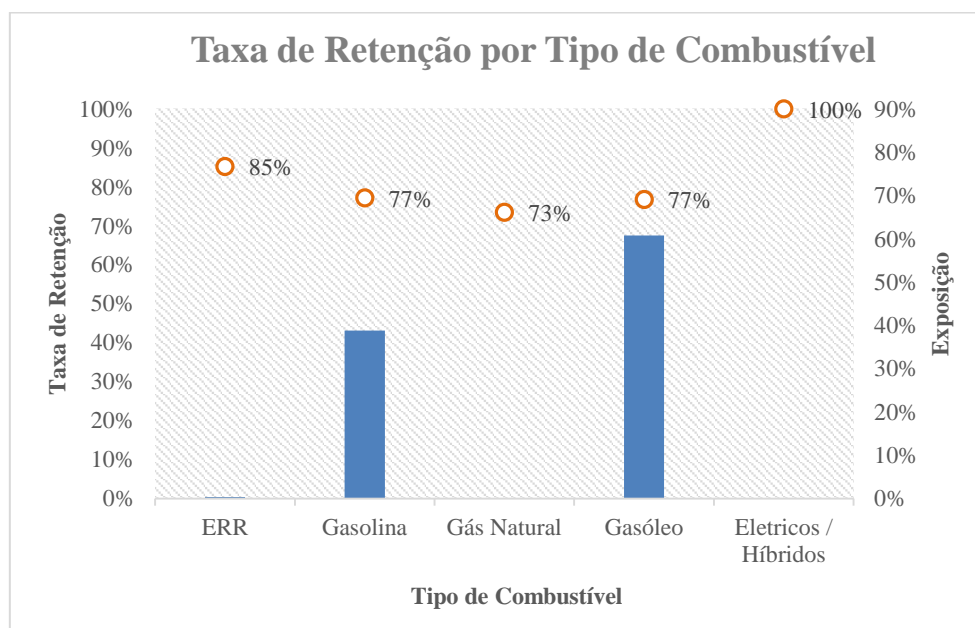


Gráfico 6.17 - Taxa de Retenção por Tipo de Combustível

A variável independente tipo de combustível é apresentada no *Gráfico 6.17*. Esta variável apresenta os quatro tipos de combustíveis possíveis dos veículos considerados. Os ERR são os veículos ao qual a seguradora não teve resposta por parte do cliente quanto ao seu tipo de combustível. Verifica-se que os veículos elétricos têm maior taxa de renovação por ainda serem muito recentes no mercado e não existir um grande número de observações. Já os veículos a gasolina e gasóleo apresentam a mesma probabilidade de renovação.

- Rácio Peso-Potência

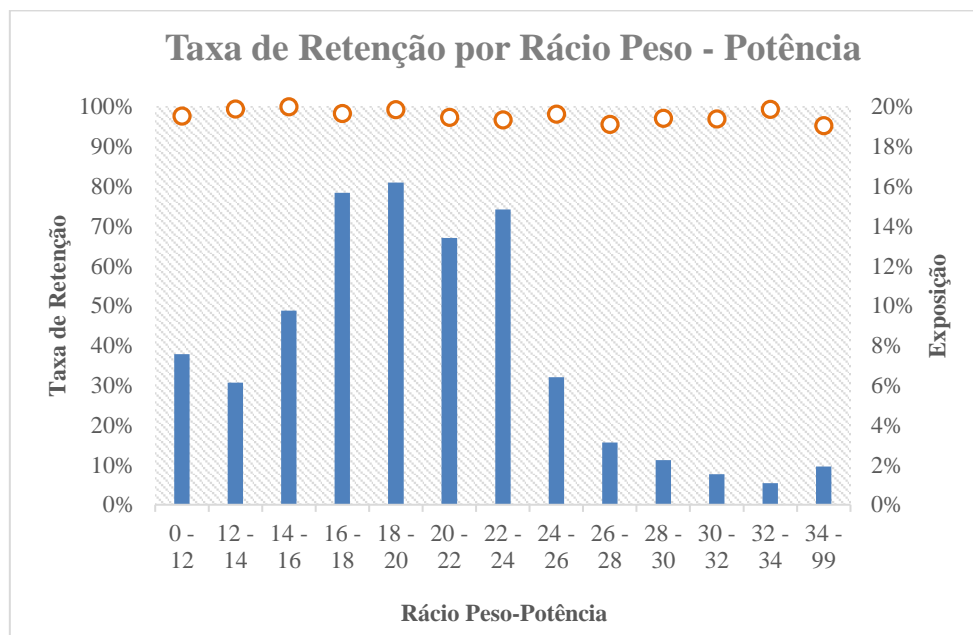


Gráfico 6.18 - Taxa de Retenção por Rácio Peso-Potência

O *Gráfico 6.18* mostra-nos a variação da taxa de retenção pelo rácio Peso – Potência. Esta variável tenta recriar um rácio entre o peso do veículo e a sua potência. É calculado através da forma $\frac{2300}{120} = 19.17$. Neste caso, o veículo tem 2300 kg e 120 cavalos. Consequentemente está na banda 18-20. Podemos observar que a taxa de retenção se mantém praticamente uniforme.

- Distrito

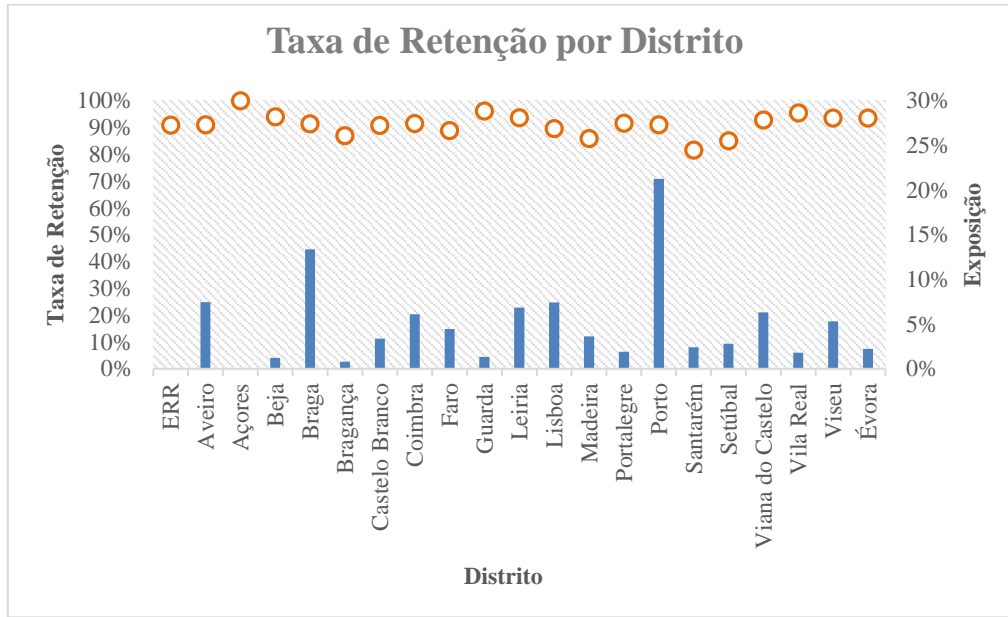


Gráfico 6.19 - Taxa de Retenção por Distrito

O Gráfico 6.19 mostra como varia a taxa de retenção pelo Distrito de circulação do Segurado. A partir da análise a este gráfico verifica-se que há um desvio considerável entre os Açores e Santarém. Já Lisboa e Porto, os distritos mais populosos do país, apresentam uma taxa de retenção idêntica.

- Valor do Veículo

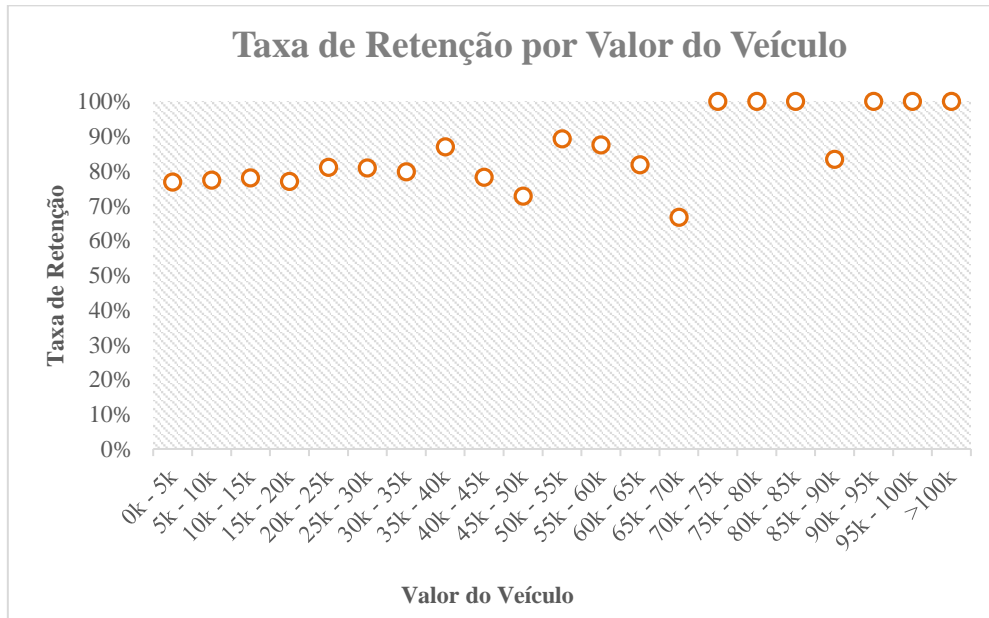


Gráfico 6.20 - Taxa de Retenção por Valor do Veículo

O Gráfico 6.20 mostra-nos como varia a taxa de retenção de acordo com o valor do veículo de acordo com a tabela de desvalorização. Para ser mais compreensível de interpretar e

analisar, criaram-se bandas para os valores dos veículos. É de notar que os veículos mais caros apresentam uma maior retenção que os restantes veículos.

- Idade do Condutor

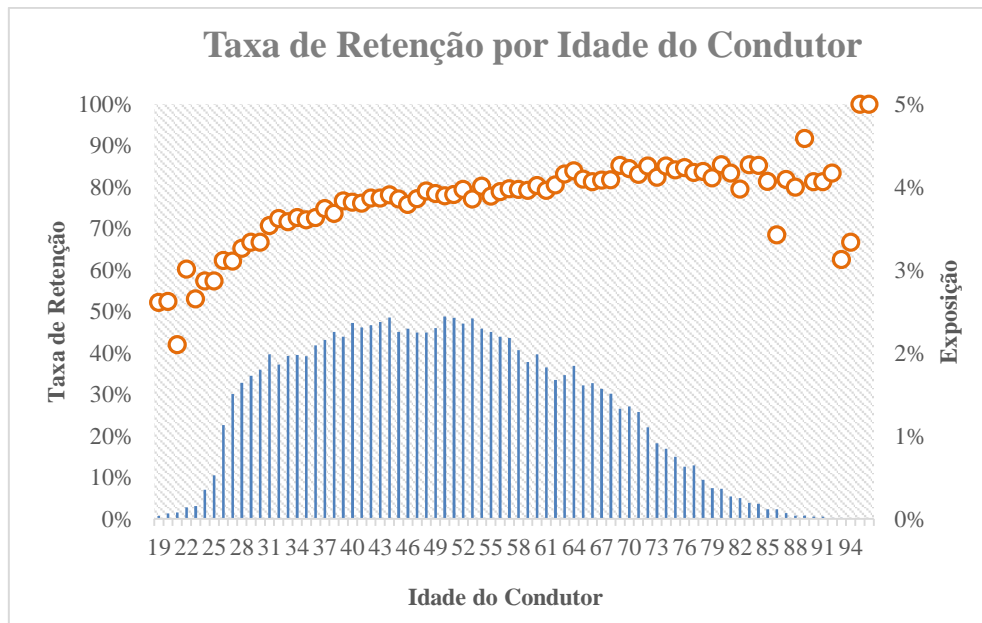


Gráfico 6.21 - Taxa de Retenção por Idade do Condutor

O Gráfico 6.21 representa a variação da taxa de retenção de acordo com a idade do condutor. É de verificar que à medida que a idade do condutor aumenta, maior é a probabilidade de renovação. Isto pode dever-se ao facto de pessoas mais novas terem prémios consideravelmente mais elevados que as pessoas mais velhas e por terem menos poder de compra. Consequentemente estão sempre à procura de preços mais competitivos no mercado segurador.

- Idade do Segurado

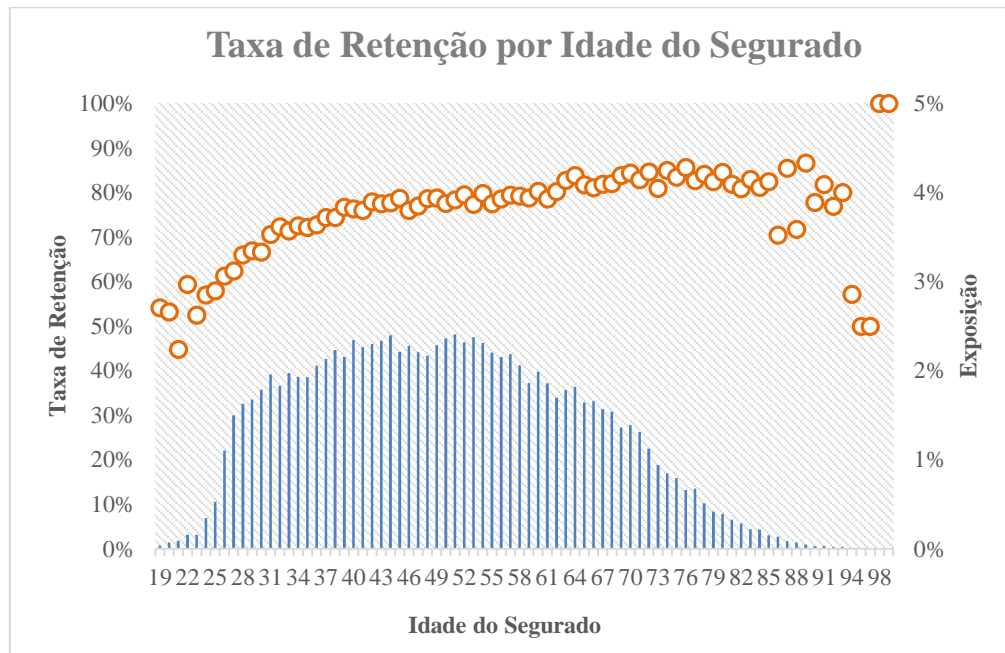


Gráfico 6.22 - Taxa de Retenção por Idade do Segurado

O Gráfico 6.22 representa a variação da taxa de retenção de acordo com a idade do segurado. A idade do segurado ajuda a definir o prêmio a ser pago pelo cliente. É possível verificar que, como no Gráfico 6.21, à medida que a idade do segurado é maior, maior é também a probabilidade de retenção.

- Idade do Veículo

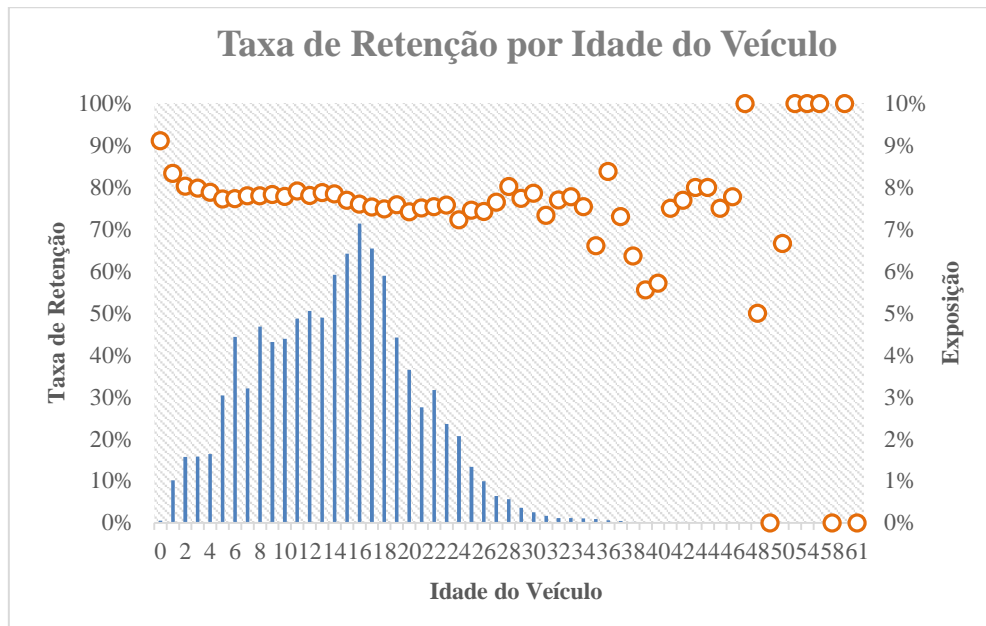


Gráfico 6.23 - Taxa de Retenção por Idade do Veículo

O Gráfico 6.23 descreve como taxa de retenção varia de acordo com a variável Idade do Veículo. Observa-se que nos primeiros 35 anos de vida do veículo, o seu comportamento se mantém estável, embora apresente uma ligeira diminuição à medida que se torna mais antigo. Esta diminuição ocorre porque os clientes têm tendência a comprar carros novos e fazerem novas apólices. Devido à pouca exposição nas apólices cuja idade do veículo é maior, obtém-se um grande ruído nos dados.

- Antiguidade da Carta de Condução

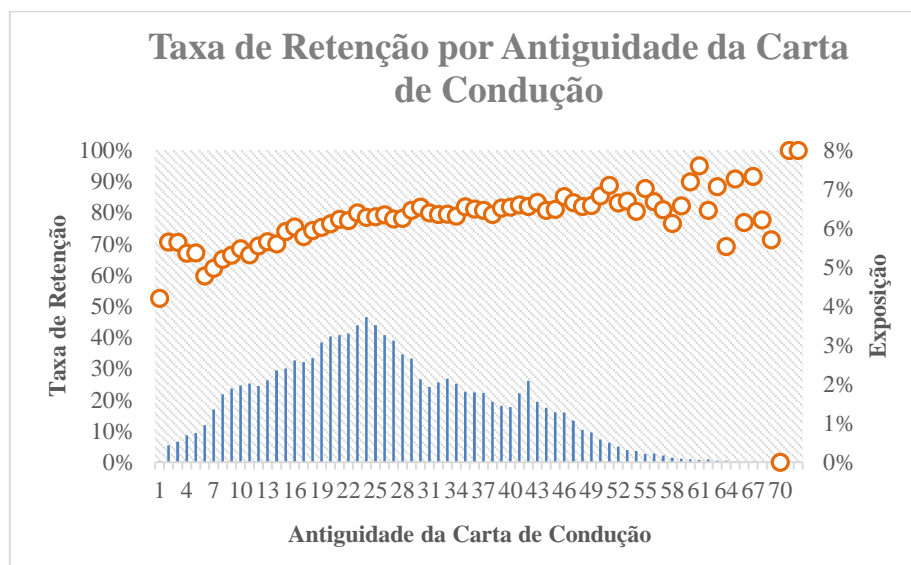


Gráfico 6.24 - Taxa de Retenção por Antiguidade da Carta de Condução

O Gráfico 6.24 demonstra como varia a taxa de retenção de acordo com o número de anos de carta de condução que o segurado tem. Esta variável define muitas vezes a experiência que o segurado tem em termos de condução. Também é verdade que quantos mais anos o cliente tem de carta de condução, maior é a sua idade. Este gráfico segue a mesma tendência dos gráficos apresentados anteriormente. Desta forma, pode-se concluir que à medida que o cliente tem mais anos de carta, maior é a sua taxa de retenção.

- Antiguidade da Apólice

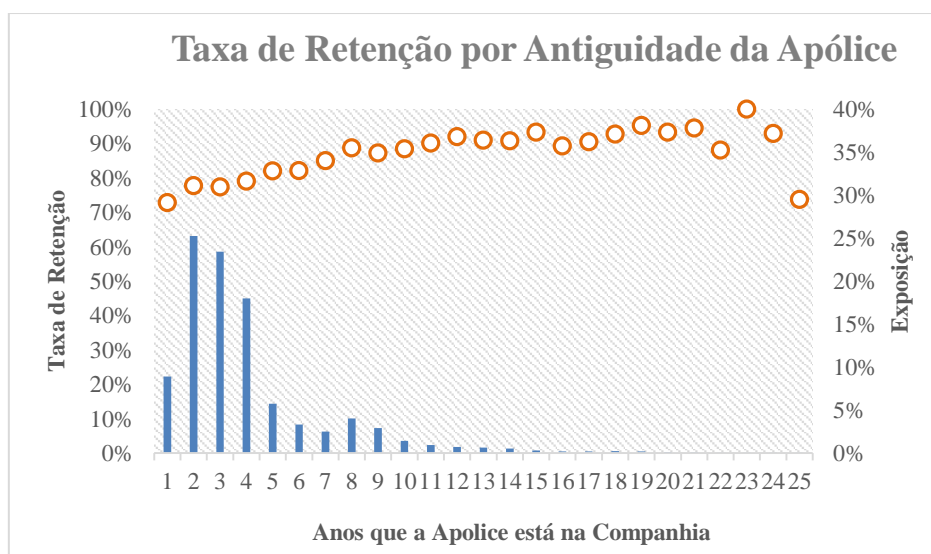


Gráfico 6.25 - Taxa de Retenção por Antiguidade da Apólice

O Gráfico 6.25 mostra-nos como varia a taxa de retenção de acordo com a Antiguidade da Apólice. Como seria de esperar, à medida que a apólice é mais antiga e está há mais anos na companhia, maior é a tendência do segurado renovar com a companhia. Sendo cliente há vários anos, e mantendo o seu grau de satisfação, muito dificilmente não irá renovar contrato.

- Retenção Média do Agente

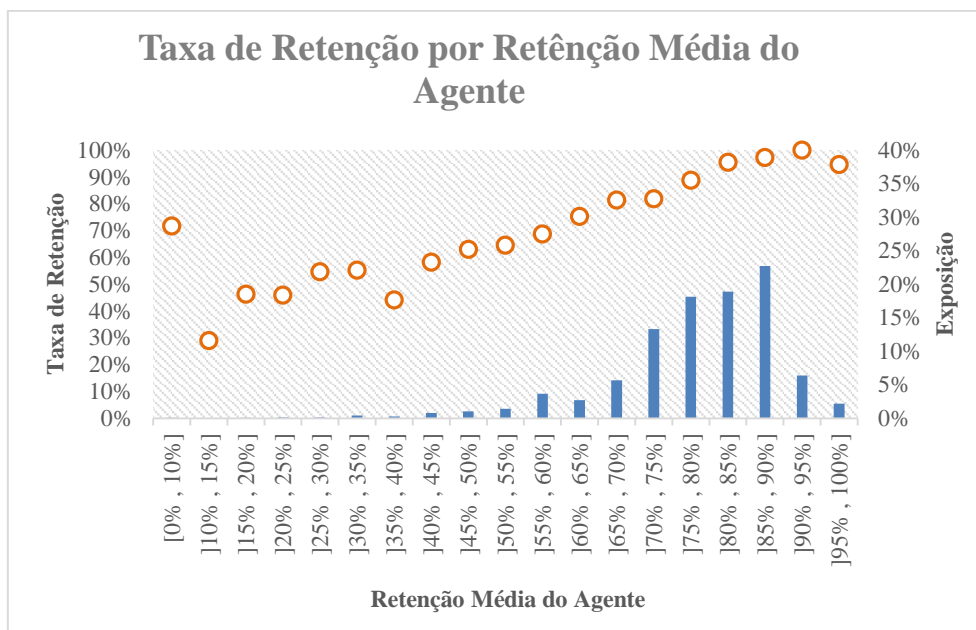


Gráfico 6.26 - Taxa de Retenção por Retenção Média do Agente

Por fim, o Gráfico 6.26 reflete a variação da taxa de retenção de acordo com a retenção média do seu agente. Este gráfico transmite a ideia que à medida que a retenção média do agente é maior, maior também é a taxa de renovação para veículos (Ramo Automóvel). Esta taxa de renovação média é para todos os ramos e não apenas para o ramo automóvel. Para a banda de [0,10%] a taxa de retenção apresentada é superior a outras por possuir algum ruído.

6.4. Seleção de Variáveis

Numa fase primária, de forma a reduzir o número de variáveis a incluir no modelo, foi-se verificar a existência de multicolinearidade das variáveis numéricas do modelo.

Utilizou-se, então, o método dos fatores de inflação da variância pois foi um método aprendido nas unidades curriculares (ver programa no anexo C) e obteve-se então o seguinte output:

Tabela 6.2 - Multicolinearidade das variáveis - Fase 1

Variável independente	VIF
Prémio Pago na anuidade de 2015	251,48
Prémio proposto na renovação	245,74
Idade do Condutor	16,08
Idade do Segurado	13,49
Antiguidade da Carta de Condução	3,78
Variação do Prémio	3,09
Valor do Veículo	2,78
Frequência de Pagamento	1,68
Idade do Veículo	1,27
Antiguidade da Apólice	1,18
Percentagem de Bónus-Malus	1,12
Rácio entre o Peso e a Potência	1,05
Retenção Média do Agente	1,03
Mês de Renovação	1,02

Dado estes resultados, conclui-se que a variável Prémio pago na anuidade de 2015 já está a ser explicada pelas outras variáveis numéricas. Assim, retira-se essa variável pois não acrescentara algo bom ao modelo.

O procedimento repetiu-se e obteve-se o seguinte output:

Tabela 6.3 - Multicolinearidade das variáveis - Fase 2

Variável independente	VIF
Idade do Condutor	16,08
Idade do Segurado	13,49
Antiguidade da Carta de Condução	3,78
Prémio proposto na renovação	2,87
Valor do Veículo	2,76
Variação do Prémio	1,77
Frequência de Pagamento	1,68
Idade do Veículo	1,27
Antiguidade da Apólice	1,18
Percentagem de Bónus-Malus	1,12
Rácio entre o Peso e a Potência	1,05
Retenção Média do Agente	1,03
Mês de Renovação	1,02

A análise VIF permitiu concluir que a idade do condutor já está a ser explicada por outras variáveis. Sendo assim, retira-se esta variável independente no modelo.

Seguidamente fez-se a mesma análise VIF para as restantes variáveis.

Tabela 6.4 - Multicolinearidade das variáveis - Fase 3

Variável independente	VIF
Idade do Segurado	3,19
Antiguidade da Carta de Condução	3,18
Prémio proposto na renovação	2,87
Valor do Veículo	2,76
Variação do Prémio	1,77
Frequência de Pagamento	1,68
Idade do Veículo	1,27
Antiguidade da Apólice	1,18
Percentagem de Bónus-Malus	1,12
Rácio Peso-Potência	1,05
Retenção Média do Agente	1,03
Mês de Renovação	1,02

Como todos os fatores da variância são inferiores a 10, significa que não existem mais variáveis que estão altamente correlacionadas. Assim pode-se partir para a criação de um modelo de regressão logística.

O método utilizado na construção do modelo foi o Backward. Começou-se com todas as variáveis no modelo e verifica-se a que apresenta maior p_value. Se o p_value desta variável for superior a 5% elimina-se do modelo. As variáveis independentes vão sendo excluídas até ficar-se apenas com variáveis explicativas. Inicialmente começamos com o modelo completo sem as variáveis Idade do condutor e prémio proposto para a antiguidade de 2015 pois já foram excluídas devido aos fortes graus de multicolinearidade.

As variáveis qualitativas foram transformadas em variáveis dummy. Segue um exemplo da transformação para a variável Grupo de Agentes:

Tabela 6.5 - Exemplo de Variável Dummy

Grupo de Agentes	Variáveis Dummy		
	D1	D2	D3
Bronze	1	0	0
Silver	-1	-1	-1
Gold	0	1	0
Platinum	0	0	1

A D1 representa a variável Dummy_Bronze, a D2 a Dummy_Gold e a D3 a Dummy_Platinum. Assim, o que antes era uma única variável qualitativa com quatro diferentes níveis, agora transformou-se em três variáveis dummy. Se uma observação tem um agente que

pertença ao grupo bronze, então o seu coeficiente irá ser 1 na D1 e 0 na D2 e D3. Contudo se for silver, então terá o coeficiente -1 em todas as dummy associadas ao grupo de agentes.

Relativamente às restantes variáveis qualitativas, a marca do veículo que apresenta valor -1 é os “Outros”, o tipo de veículo é os “Auto-Caravanas”, o distrito é “Evora”, o tipo de combustível é os “elétricos” e a cobertura da franquia de danos próprios é o “Not Applicable”.

A tabela seguinte apresenta todas as iterações realizadas de exclusão de variáveis. Apresenta-se então por ordem a variável que foi excluída e o seu p_value. O anexo D reproduz o programa utilizado:

Tabela 6.6 - Variáveis a eliminar do modelo

Passo	Variável a eliminar	Graus de Liberdade	Wald Chi-Square	Probabilidade > ChiSQ
1	Dummy_Rover	1	0.0025	0.9604
2	Dummy_Franquia_j(DP)	1	0.0036	0.9519
3	Dummy_Elétricos	1	0.0006	0.9806
4	Dummy_Nissan	1	0.0092	0.9238
5	Dummy_Franquia_h(DP)	1	0.0103	0.9193
6	Dummy_Mazda	1	0.0194	0.8892
7	Dummy_Ligeiros	1	0.0321	0.8578
8	Racio_Peso_Potência	1	0.0314	0.8593
9	Dummy_Açores	1	0.0429	0.8359
10	Dummy_Kia	1	0.0675	0.7950
11	Dummy_TTLigeiros	1	0.0778	0.7804
12	Tem_Quebra_Vidros	1	0.0846	0.7711
13	dummy_Iveco	1	0.1021	0.7493
14	Dummy_Peugeot	1	0.1233	0.7255
15	Dummy_Seat	1	0.1624	0.6870
16	Zona_Geográfica	1	0.2447	0.6208
17	Dummy_Portalegre	1	0.1835	0.6684
18	Dummy_Mini	1	0.2771	0.5986
19	Dummy_TTMisto	1	0.3011	0.5832
20	Dummy_LandRover	1	0.2335	0.6290
21	Dummy_Citroen	1	0.2687	0.6042
22	Dummy_Skoda	1	0.2924	0.5887
23	Dummy_Franquia_e(DP)	1	0.3317	0.5647
24	Dummy_Franquia_l(DP)	1	0.5770	0.4475
25	Dummy_Volvo	1	0.6102	0.4347
26	Dummy_Misto	1	0.5815	0.4457
27	Dummy_Gasolina	1	0.6878	0.4069

28	Dummy_Gasóleo	1	0.3007	0.5835
29	Dummy_Franquia_k(DP)	1	0.7215	0.3957
30	Dummy_Volswagen	1	0.7766	0.3782
31	Dummy_Franquia_a(DP)	1	0.7877	0.3748
32	Dummy_Erros(Distrito)	1	0.7979	0.3717
33	Dummy_Viseu	1	0.5282	0.4673
34	Dummy_Hyundai	1	11.823	0.2769
35	Dummy_Audi	1	13.067	0.2530
36	Dummy_Lancia	1	13.107	0.2523
37	Dummy_BMW	1	11.472	0.2841
38	Dummy_Lisboa	1	13.545	0.2445
39	Dummy_Gás_Natural	1	14.752	0.2245
40	Dummy_Franquia_b(DP)	1	16.699	0.1963
41	Dummy_Leiria	1	16.874	0.1939
42	Dummy_Franquia_n(DP)	1	17.107	0.1909
43	Dummy_Honda	1	19.801	0.1594
44	Dummy_Coimbra	1	19.860	0.1588
45	Dummy_Porto	1	18.070	0.1789
46	Dummy_Faro	1	20.608	0.1511
47	Dummy_Mercedes	1	20.984	0.1475
48	Dummy_i_Franquia	1	24.559	0.1171
49	Dummy_Opel	1	25.480	0.1104
50	Dummy_Setúbal	1	27.410	0.0978
51	Dummy_Franquia_d(DP)	1	30.877	0.0789
52	Dummy_Franquia_g(DP)	1	27.657	0.0963
53	Dummy_Suzuki	1	31.435	0.0762
54	Dummy_Chevrolet	1	24.046	0.1210

Após a exclusão da variável dummy referente aos veículos de marca Chevrolet, mais nenhuma variável era candidata a sair. Sendo assim obteve-se um modelo de regressão logística com as variáveis que estão representadas na tabela seguinte:

Tabela 6.7 - Variáveis selecionadas para o modelo

Variáveis que ficam no modelo	Graus de Liberdade	Betas Estimados	Standard Error	Wald Chi-Square	Probabilidade > ChiSQ
β_0	1	-0.8851	0.1134	609.206	<.0001
Tipo_Cliente	1	0.0863	0.0222	150.587	0.0001
Dummy_Bronze	1	-0.1350	0.0273	244.027	<.0001
Dummy_Gold	1	0.0653	0.0172	143.398	0.0002
Dummy_Platinum	1	0.1903	0.0255	557.920	<.0001
Mês de Renovação	1	0.0521	0.00435	1.437.871	<.0001
Dummy_Caminheta	1	-0.2054	0.0496	171.510	<.0001
Dummy_ttcamineta	1	0.2772	0.1062	68.120	0.0091
Dummy_Comercial	1	-0.1395	0.0360	149.834	0.0001
Tem_DanosProprios	1	0.4675	0.0504	860.712	<.0001
Bonus_Malus	1	0.0223	0.00176	1.607.571	<.0001
Prémio Proposto na renovação	1	-0.00248	0.00015	2.569.893	<.0001
Variação do Prémio	1	-0.0798	0.00284	7.897.933	<.0001
Dummy_Renault	1	-0.0723	0.0301	57.643	0.0164
Dummy_Ford	1	-0.0831	0.0395	44.201	0.0355
Dummy_Fiat	1	-0.0917	0.0434	44.580	0.0347
Dummy_Toyota	1	0.2122	0.0453	219.820	<.0001
Dummy_Mitsubishi	1	0.1705	0.0621	75.360	0.0060
Dummy_Smart	1	-0.3129	0.0932	112.672	0.0008
Dummy_AlfaRomeo	1	-0.3026	0.1138	70.673	0.0079
Dummy_GamaAlta	1	0.5032	0.1510	111.090	0.0009
Dummy_Aveiro	1	-0.1019	0.0366	77.645	0.0053
Dummy_Beja	1	0.1861	0.0864	46.421	0.0312
Dummy_Braga	1	0.1956	0.0292	449.929	<.0001
Dummy_Bragança	1	-0.3273	0.0976	112.459	0.0008
Dummy_Castelo Branco	1	-0.1166	0.0520	50.302	0.0249
Dummy_Guarda	1	0.3490	0.0872	160.288	<.0001
Dummy_Madeira	1	-0.3560	0.0490	527.841	<.0001
Dummy_Santarém	1	-0.2275	0.0562	163.904	<.0001
Dummy_Viana do Castelo	1	0.1078	0.0410	69.151	0.0085
Dummy_VilaReal	1	0.3124	0.0745	175.675	<.0001
Valor do Veículo	1	0.000032	0.00339	891.021	<.0001
Idade do Segurado	1	0.00990	0.00122	658.764	<.0001
Idade do Veículo	1	-0.0219	0.00181	1.466.090	<.0001
Antiguidade da Carta de Condução	1	0.00325	0.00145	50.575	0.0245
Antiguidade da Apólice	1	0.0300	0.00414	525.926	<.0001
Frequência de Pagamento	1	0.1272	0.0157	660.045	<.0001
Retenção Média do Agente	1	0.0146	0.00057	6.497.446	<.0001

O modelo criado é da seguinte forma:

$$\begin{aligned} \text{Seja } LP = & - 0.8851 + 0.0863 * \text{Tipo_Cliente} - 0.1350 * \text{Dummy_Bronze} + \\ & 0.0653 * \text{Dummy_Gold} + 0.1903 * \text{Dummy_Platinum} + 0.0521 * \text{Mês_Renovação} - \\ & 0.2054 * \text{Dummy_Caminheta} + 0.2772 * \text{Dummy_ttCaminheta} - 0.1395 * \text{Dummy_Comercial} + \\ & 0.4675 * \text{Tem_DanosProprios} + 0.0223 * \text{Bonus_Malus} - 0.00248 * \text{Premio_Proposto} - \\ & 0.0798 * \text{Variação_Premio} - 0.0723 * \text{Dummy_Renault} - 0.0831 * \text{Dummy_Ford} - \\ & 0.0917 * \text{Dummy_Fiat} + 0.2122 * \text{Dummy_Toyota} + 0.1705 * \text{Dummy_Mitsubishi} - \\ & 0.3129 * \text{Dummy_Smart} - 0.3026 * \text{Dummy_AlfaRomeu} + 0.5032 * \text{Dummy_GamaAlta} - \\ & 0.1019 * \text{Dummy_Aveiro} + 0.1861 * \text{Dummy_Beja} + 0.1956 * \text{Dummy_Braga} - \\ & 0.3273 * \text{Dummy_Bragança} - 0.1166 * \text{Dummy_CasteloBranco} + 0.319 * \text{Dummy_Guarda} - \\ & 0.3560 * \text{Dummy_Madeira} - 0.2275 * \text{Dummy_Santarem} + 0.1078 * \text{Dummy_VianadoCastelo} + \\ & 0.3124 * \text{Dummy_VilaReal} + 0.000032 * \text{Valor_Veículo} + 0.0099 * \text{Idade_Segurado} - \\ & 0.0219 * \text{IdadeVeículo} + 0.00325 * \text{AnosCartaCondução} + 0.03 * \text{Antiguidade_Apolice} + \\ & 0.1272 * \text{Frequencia_Pagamento} + 0.0146 * \text{Retenção_Media_Agente}. \end{aligned}$$

A probabilidade de retenção da apólice i é dado por $\frac{1}{1+\exp(-LP)}$

As variáveis quantitativas são representadas pelo seu valor e as restantes como tendo o valor de:

- Tipo do cliente
 - 1 se o cliente for de alto valor
 - 0 se o cliente pertencer ao grupo restantes
- Dummy_Bronze
 - 1 se o agente da apólice for do tipo Bronze
 - 0 se o agente da apólice não for do tipo Bronze ou Silver
 - -1 se o agente da apólice for do tipo Silver
- Dummy_Gold
 - 1 se o agente da apólice for do tipo Gold
 - 0 se o agente da apólice não for do tipo Gold ou Silver
 - -1 se o agente da apólice for do tipo Silver
- Dummy_Platinum
 - 1 se o agente da apólice for do tipo Platinum
 - 0 se o agente da apólice não for do tipo Platinum ou Silver
 - -1 se o agente da apólice for do tipo Silver
- Dummy_Caminheta
 - 1 se o veículo seguro for uma caminheta
 - 0 se o veículo seguro não for Caminheta ou autocaravana
 - -1 se o veículo seguro for uma Autocaravana
- Dummy_ttCaminheta
 - 1 se o veículo seguro for uma tt caminheta
 - 0 se o veículo seguro não for tt Caminheta ou autocaravana
 - -1 se o veículo seguro for uma Autocaravana

- Dummy_Comercial
 - 1 se o veículo seguro for um comercial
 - 0 se o veículo seguro não for comercial ou autocaravana
 - -1 se o veículo seguro for uma Autocaravana
- Tem_DanosProprios
 - 1 se a observação tem a cobertura de danos próprios
 - 0 se a observação não tem a cobertura de danos próprios
- Dummy_Renault
 - 1 se o veículo seguro for da marca Renault
 - 0 caso o veículo seguro não seja da marca Renault ou “outros”
 - -1 se o veículo seguro for da marca “outros”
- Dummy_Ford
 - 1 se o veículo seguro for da marca Ford
 - 0 caso o veículo seguro não seja da marca Ford ou “outros”
 - -1 se o veículo seguro for da marca “outros”
- Dummy_Fiat
 - 1 se o veículo seguro for da marca Fiat
 - 0 caso o veículo seguro não seja da marca Fiat ou “outros”
 - -1 se o veículo seguro for da marca “outros”
- Dummy_Toyota
 - 1 se o veículo seguro for da marca Toyota
 - 0 caso o veículo seguro não seja da marca Toyota ou “outros”
 - -1 se o veículo seguro for da marca “outros”
- Dummy_Mitsubishi
 - 1 se o veículo seguro for da marca Mitsubishi
 - 0 caso o veículo seguro não seja da marca Mitsubishi ou “outros”
 - -1 se o veículo seguro for da marca “outros”
- Dummy_Smart
 - 1 se o veículo seguro for da marca Smart
 - 0 caso o veículo seguro não seja da marca Smart ou “outros”
 - -1 se o veículo seguro for da marca “outros”
- Dummy_AlfaRomeu
 - 1 se o veículo seguro for da marca Alfa Romeu
 - 0 caso o veículo seguro não seja da marca Alfa Romeu ou “outros”
 - -1 se o veículo seguro for da marca “outros”
- Dummy_GamaAlta
 - 1 se o veículo seguro for da marca Gama Alta
 - 0 caso o veículo seguro não seja da marca Gama Alta ou “outros”
 - -1 se o veículo seguro for da marca “outros”
- Dummy_Aveiro
 - 1 se o distrito de circulação do cliente for Aveiro
 - 0 caso o distrito de circulação não seja Aveiro ou Évora
 - -1 se o distrito de circulação do cliente for Évora

- Dummy_Beja
 - 1 se o distrito de circulação do cliente for Beja
 - 0 caso o distrito de circulação não seja Beja ou Évora
 - -1 se o distrito de circulação do cliente for Évora
- Dummy_Braga
 - 1 se o distrito de circulação do cliente for Braga
 - 0 caso o distrito de circulação não seja Braga ou Évora
 - -1 se o distrito de circulação do cliente for Évora
- Dummy_Bragança
 - 1 se o distrito de circulação do cliente for Bragança
 - 0 caso o distrito de circulação não seja Bragança ou Évora
 - -1 se o distrito de circulação do cliente for Évora
- Dummy_CasteloBranco
 - 1 se o distrito de circulação do cliente for Castelo Branco
 - 0 caso o distrito de circulação não seja Castelo Branco ou Évora
 - -1 se o distrito de circulação do cliente for Évora
- Dummy_Guarda
 - 1 se o distrito de circulação do cliente for Guarda
 - 0 caso o distrito de circulação não seja Guarda ou Évora
 - -1 se o distrito de circulação do cliente for Évora
- Dummy_Madeira
 - 1 se o distrito de circulação do cliente for Madeira
 - 0 caso o distrito de circulação não seja Madeira ou Évora
 - -1 se a zona de circulação do cliente for Évora
- Dummy_Santarém
 - 1 se o distrito de circulação do cliente for Santarém
 - 0 caso o distrito de circulação não seja Vila Real ou Évora
 - -1 se a zona de circulação do cliente for Évora
- Dummy_VianaDoCastelo
 - 1 se o distrito de circulação do cliente for Viana do Castelo
 - 0 caso o distrito de circulação não seja Vila Real ou Évora
 - -1 se o distrito de circulação do cliente for Évora
- Dummy_VilaReal
 - 1 o distrito de circulação do cliente for Vila Real
 - 0 caso o distrito de circulação não seja Vila Real ou Évora
 - -1 se o distrito de circulação do cliente for Évora
- Frequência de Pagamento
 - 1 se a forma de pagamento for anual
 - 2 se a forma de pagamento for semestral
 - 4 se a forma de pagamento for trimestral

Um individuo que tenha 71 anos de idade, circule no distrito de lisboa, que tirou a carta há 49 anos, seja cliente da companhia há 8 anos, seja um cliente de alto valor, cujo agente tenha uma retenção média de 87.5 e que pertença ao grupo Gold, que renove a sua apólice no mês 5, tenha a cobertura de danos próprios, que tenha uma letra de bónus-malus J, ao qual o prémio proposto seja 705€, e sofra um aumento de 7,5%, cujo seu veículo seja um Volvo , tenha 9 anos

e um valor de 7500€, e que a forma de pagamento do prémio seja anual apresenta um LP de 1.08935, e consequentemente uma percentagem de retenção de 74.82%.

6.5. Diagnóstico do Modelo

Após análise da tabela anterior, conclui-se que o modelo de regressão logística apresenta 37 variáveis.

Constata-se que:

-Todas as dummy criadas para o Grupo do Agente são importantes, sendo que apenas os Agentes Bronze apresentam um beta negativo (-0.1350). Isto significa que se a apólice tiver um agente Bronze, tem uma menor taxa de retenção que as restantes apólices que possuem outro tipo de agente e as mesmas características nas restantes variáveis independentes. Relativamente aos Agentes Silver, o seu beta é de $(-0.1350+0.0653+0.1903) \cdot -1$, ou seja, -0.1206, pelo que apresenta um beta negativo. Estes valores vão ao encontro do que se via no gráfico relativo ao Grupo de Agentes. As restantes variáveis dummy do grupo de agentes apresentam um beta positivo;

-Todas as variáveis dummy criadas para a franquia da Cobertura de Danos próprios foram eliminadas, ou seja, esta variável não é importante para o modelo;

- O mês de renovação é uma das variáveis explicativas, apresentando um beta de 0.0521. Ou seja, quanto maior for o mês de renovação, maior será a sua probabilidade de retenção (não esquecendo que se tem de manter as restantes características);

- As variáveis independentes “Tipo de Cliente”, “Existência de Danos Próprios”, “Bónus-Malus”, “Valor do Veículo”, “Idade do Segurado”, “Antiguidade da carta de condução”, “Antiguidade da apólice”, “Frequência de Pagamento” e “Retenção Média do Agente” apresentam betas positivos;

- As variáveis independentes “prémio proposto”, “variação do prémio” e “idade do veículo” apresentam betas negativos. Isto significa que quanto maior for o valor de qualquer uma destas variáveis, menor será a sua taxa de retenção (supondo que tem as restantes variáveis iguais);

-As variáveis que estão no modelo referentes ao tipo de veículo são as “Caminhetas”, “TTCaminhetas” e “Comercial”. Apenas a “TTCaminheta” apresenta um beta positivo;

-As variáveis importantes referentes à marca do veículo para a modelização da taxa de retenção são a “Renault”, “Ford”, “Fiat”, “Toyota”, “Mitsubishi”, “Smart”, “Alfa Romeu” e os “veículos de Gama Alta”. É de referir que as apólices que apresentem carros das seguintes marcas: “Renault”, “Ford”, “Fiat”, “Smart” e “Alfa Romeu” apresentam uma menor taxa de retenção que as restantes apólices que apresentem as mesmas características (ao possuírem betas negativos). De forma análoga ao cálculo do beta para a variável “Agentes Silver”, verifica-se que a marca referente a “Outros”, apresenta um beta de -0.023;

- Os distritos que entram neste modelo de Regressão Logística são “Aveiro”, “Beja”, “Braga”, “Bragança”, “Castelo Branco”, “Guarda”, “Madeira”, “Santarém”, “Viana do Castelo” e “Vila Real”, sendo que “Aveiro”, “Bragança”, “Castelo Branco”, “Madeira”, e “Santarém” fazem com

que a apólice obtenha uma menor taxa de retenção. As restantes variáveis apresentam betas positivos. “Evora” apresenta um beta de -0.0216;

-As variáveis correspondentes à zona geográfica, existência da cobertura de quebra isolada de vidros, a franquia correspondente à cobertura de danos próprios, prémio pago no ano de 2015, tipo de combustível, rácio peso-potência e idade do condutor não estão presentes no modelo construído.

6.5.1. Matriz de Confusão e Curva de ROC

Tal como descrito nos capítulos anteriores, uma forma prática de analisar os resultados obtidos é através de uma matriz de confusão. Iremos construir duas tabelas de classificação distintas para dois pontos de corte. Assim, o modelo anterior possui a seguinte tabela para os diferentes pontos de corte:

Tabela 6.8 - Diferentes pontos de corte para o modelo criado

Pontos de Corte	1 - Especificidade	Sensibilidade	Especificidade
0,2	100%	100%	0%
0,3	99%	100%	1%
0,4	97%	99%	3%
0,5	91%	98%	9%
0,6	80%	94%	20%
0,7	59%	83%	41%
0,8	29%	56%	71%
0,85	14%	32%	86%
0,9	3%	9%	97%
0,95	0%	0%	100%

A partir do mesmo podemos construir o gráfico que nos indica qual o ponto de corte perfeito:

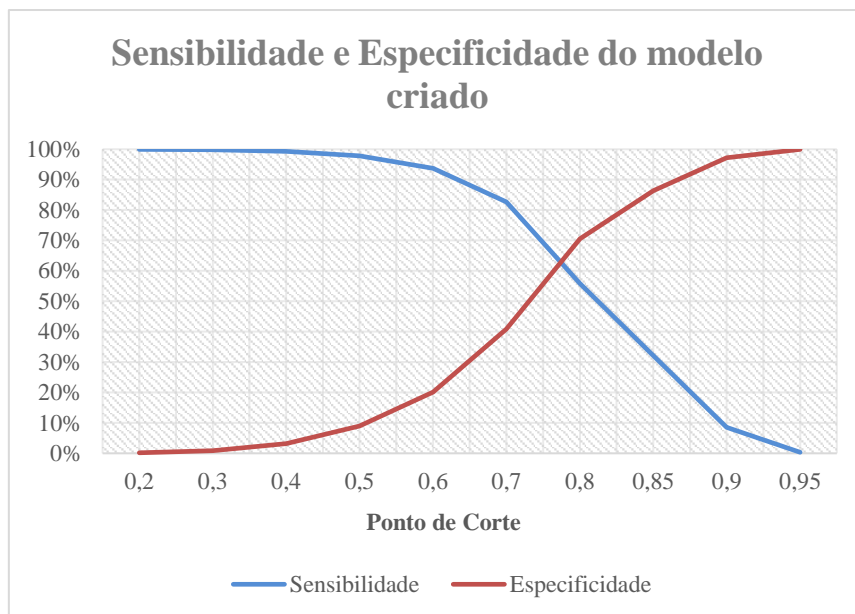


Gráfico 6.27 - Sensibilidade e Especificidade do modelo criado

Através da análise do gráfico anterior, verifica-se que o ponto de corte perfeito é de 77,9%. A partir deste ponto de Corte obtemos a seguinte Matriz de confusão:

Tabela 6.9 - Matriz de confusão do modelo criado para o seu ponto de corte perfeito

		Observados		
		Positivos	Negativos	Total
Previstos	Positivos	31376	5332	36708
	Negativos	17953	9421	27374
	Total	49329	14753	64082

Desta matriz de confusão pode-se ressaltar que $\frac{(31376+9421)}{64082} = 63.6\%$ das apólices estão bem classificadas.

Da mesma forma obtém-se a tabela para um ponto de corte de 78.38% (valor do ponto de corte perfeito do modelo da Generali que será visto no capítulo seguinte).

Tabela 6.10 - Matriz de confusão do modelo criado para o ponto de corte de 78.38%

		Observados		
		Positivos	Negativos	Total
Previstos	Positivos	30502	5101	35603
	Negativos	18827	9652	28479
	Total	49329	14753	64082

Desta repara-se que 62.6% das apólices estão bem classificadas, contudo para este ponto de corte os valores de especificidade e sensibilidade são 65.4% e 61.8%, respetivamente.

A partir desta análise podemos então construir a curva ROC do modelo:

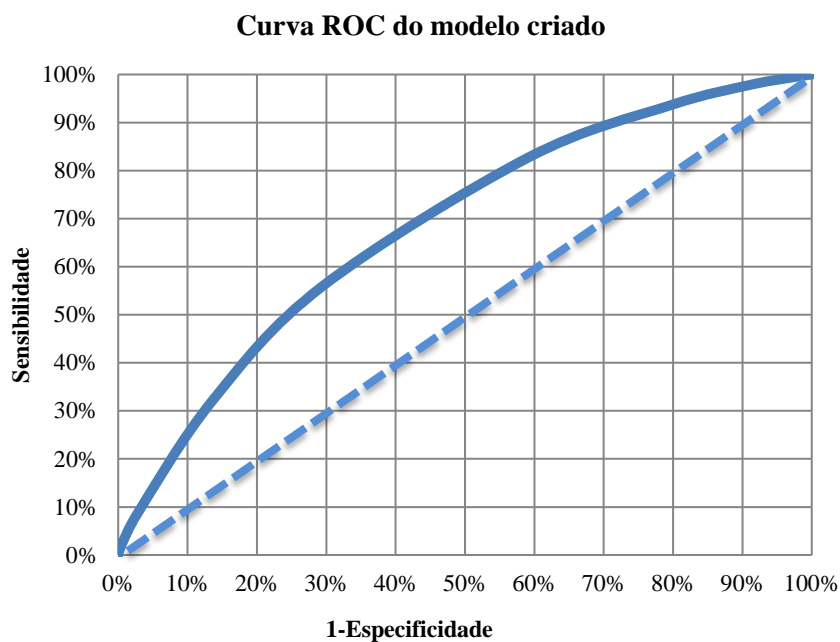


Gráfico 6.28 - Curva ROC do modelo criado

É de referir que o modelo anterior apresenta um valor de 68% de AUC, ou seja, a área abaixo da curva ROC representa 68% do gráfico apresentado. De acordo com os conceitos teóricos apresentados anteriormente, podemos dizer que o modelo criado tem uma discriminação fraca.

6.5.2. Indicadores do modelo construído

Obtendo a classificação de discriminação fraca, há outros critérios que se pode usar para posteriormente comparar com outros modelos. O modelo descrito neste capítulo apresenta as seguintes características:

- AIC = 64.589;
- AICc = 64.589;
- BIC = 69.186

De seguida vai-se ver como foi construído o modelo usado na Generali, as suas variáveis explicativas e algumas características do mesmo.

7. Modelo Generali

A companhia de seguros Generali tem desenvolvido um modelo de retenção em automóvel, que estima a probabilidade da renovação de uma apólice para um determinado perfil de risco, analisando assim o comportamento humano no momento de renovação e identificando padrões e grupos homogêneos.

O modelo existente é um modelo que tem contribuições de 3 vertentes distintas. Por um lado, é um modelo de regressão logística pois a variável resposta apresenta valores “0” ou “1”. Por outro lado, é um modelo empírico pois a maior parte das variáveis introduzidas foram colocadas através do método, tentativa erro. Por fim é um modelo que possui uma quota de experiência. Ou seja, com o trabalho conjunto da Subscrição Automóvel e Actuariado Pricing foi possível perceber através da experiência de mercado e histórico da Companhia a seleção de possíveis variáveis a serem testadas e possivelmente significativas para o modelo.

Foi criado este modelo através do software Emblem. O Emblem é um software desenvolvido pela consultora Willis Towers Watson que dá suporte técnico/estatístico a um departamento de Pricing alinhando com as suas necessidades. Na área de Actuariado Pricing da Generali é utilizado também para a criação de modelos risco (frequência e custo médio) para o Ramo Não Vida.

É um software bastante intuitivo, apresentando gráficos e um conjunto de métricas estatísticas que ajudam o utilizador a perceber como varia a distribuição dos dados e os resultados, poupando ao utilizador a criação dos mesmos através de outro software ou outras análises. Tem também a vantagem de poder suportar milhares de observações com os cálculos requeridos para a construção do modelo.

Através da possibilidade de inclusão e exclusão de variáveis, consegue-se perceber através de indicativos como o AIC, BIC ou coeficiente de Gini qual o melhor modelo.

Por exemplo, suponhamos que:

$$\beta_0 = 1.6118$$

Tabela 7.1 – Output Emblem

Prémio Proposto	Coeficiente	Forma de Pagamento	Coeficiente	Retenção Média Agente	Coeficiente
[120 , 130)	0.1580	Anual	0	25%	-2.7359
[130 , 140)	0.1264	Semestral	0,2353	30%	-2.5254
[140 , 150)	0.0948	Trimestral	0,2179	35%	-2.3150

Então, uma apólice que tenha um prémio proposto entre 120 e 129.99, com pagamento semestral e o seu agente tenha uma retenção média de 30% tem a seguinte probabilidade de retenção:

$$\text{Linear predictor (LP)} = \beta_0 + 0.1518 + 0.2353 - 2.5254 = 0.5265$$

Logo a probabilidade de retenção é $\frac{\exp(LP)}{1+\exp(LP)} = \frac{1}{1+\exp(-LP)} = 37.1\%$

Dentro do modelo da Generali estão presentes as seguintes variáveis explicativas:

- Zona Geográfica;
- Tipo de Cliente;
- Grupo de Agentes;
- Mês de Renovação;
- Existência da cobertura de quebra isolada de vidros;
- Indicador se a apólice tem a cobertura de danos próprios;
- Prémio proposto para a renovação;
- Variação entre o prémio proposto e o prémio pago em 2015;
- Distrito;
- Idade do segurado;
- Idade do Veículo;
- Antiguidade da carta de condução;
- Antiguidade da apólice;
- Frequência de Pagamento;
- Retenção média do agente.

Este modelo de retenção apresenta então as seguintes características:

- 15 variáveis independentes;
- AIC =63.575;
- AICc=63.575;
- BIC=63.929.

De forma análoga ao capítulo anterior e considerando os mesmos pontos de corte obtemos os seguintes outputs:

Tabela 7.2 - Diferentes pontos de corte para o modelo Generali

Pontos de Corte	1- Especificidade	Sensibilidade	Especificidade
0,2	99%	100%	1%
0,3	97%	100%	3%
0,4	94%	99%	6%
0,5	87%	97%	13%
0,6	76%	93%	24%
0,7	58%	84%	42%
0,8	30%	59%	70%
0,85	15%	36%	85%
0,9	3%	10%	97%
0,95	0%	0%	100%

Pode-se então construir o gráfico que nos vai ajudar a encontrar o ponto de corte perfeito. Sendo assim:

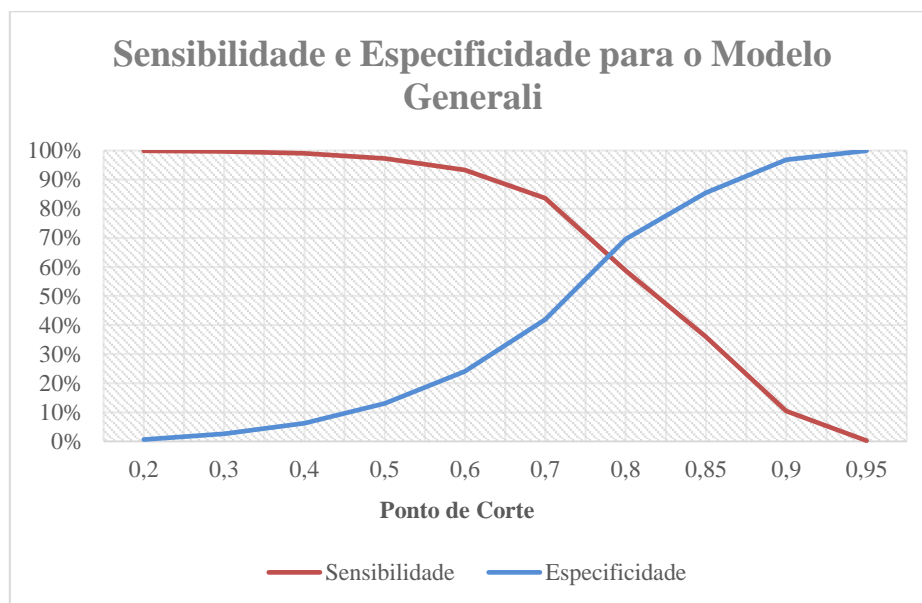


Gráfico 7.1 - Sensibilidade e Especificidade do Modelo Generali

De forma análoga ao capítulo anterior, para um ponto de corte de 77.9% (ponto de corte perfeito para o modelo construído neste trabalho), a matriz de confusão resultante do modelo da Generali é a seguinte:

Tabela 7.3 – Matriz de confusão para o ponto de corte de 77.9% referente ao modelo Generali

		Observados		Total
		Positivos	Negativos	
Previstos	Positivos	32548	5484	38032
	Negativos	16781	9269	26050
Total		49329	14753	64082

Desta matriz de confusão pode-se ressaltar que 65.25% das apólices estão bem classificadas, contudo neste caso os valores de especificidade e sensibilidade são de 62.8% e 65.98% respetivamente.

Para um ponto de corte mais alto, 78.38% (valor de corte perfeito para o modelo Generali), obtém-se a matriz de confusão seguinte:

Tabela 7.4 – Matriz de confusão para o ponto de corte perfeito do modelo Generali

		Observados		Total
		Positivos	Negativos	
Previstos	Positivos	31789	5246	37035
	Negativos	17540	9507	27047
	Total	49329	14753	64082

Desta matriz de confusão pode-se ressaltar que 64.4% das apólices estão bem classificadas, obtendo assim o mesmo valor (64.4%) de especificidade e sensibilidade.

Após esta análise e com auxílio da *Tabela 7.2* obtemos a curva ROC para este modelo.

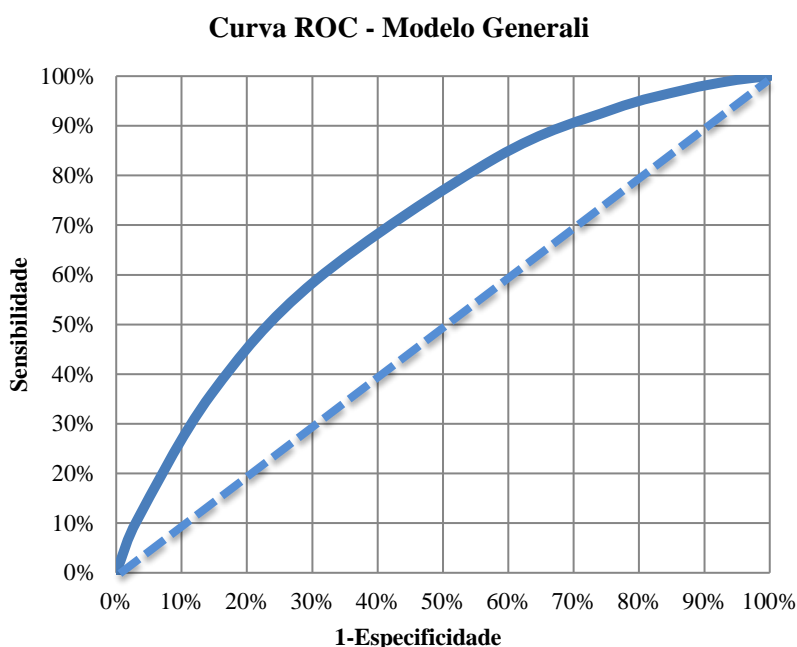


Gráfico 7.2 - Curva ROC do modelo da Generali

A partir da construção do *Gráfico 7.2* é possível referir que o valor de AUC do gráfico é de 68.7%. Este modelo possui também uma discriminação fraca.

8. Conclusões

Com a competitividade do mercado e o desenvolvimento da área de Data Analytics torna-se cada vez mais importante as seguradoras encontrarem modelos sofisticados à tomada de decisão. Ao mesmo tempo é cada vez mais importante o estudo e o foco no portfolio a renovar dado ser este o que tem maior peso no portfolio total e onde medidas de RePricing têm um maior contributo ao longo do ano para atingir os resultados de Plano e Forecast definidos.

Sabe-se que os consumidores são sensíveis à alteração de preço e daí a importância de se conseguir prever o comportamento individual à mudança de preço. Neste caso específico pretende-se prever se o cliente vai aceitar o novo prémio proposto para a renovação seguinte da sua apólice.

Um modelo de retenção para o ramo automóvel é um modelo recente no mercado português, estando a ser iniciado o seu desenvolvimento em mais companhias, seja numa abordagem mais sofisticada como a apresentada, seja numa vertente mais simples com Árvores de Decisão. Este, por si só, indica-nos qual a probabilidade de o cliente renovar a apólice para a anuidade seguinte. Esta probabilidade é distinta entre clientes, uma vez que reflete o perfil de cada um, tendo em conta as variáveis que estão incluídas no modelo.

O modelo de retenção ganha ainda mais importância quando a companhia de seguros pretende otimizar os preços das renovações. Na otimização de preços, define-se qual ou quais as variáveis a otimizar. A empresa pode, por exemplo, querer melhorar o resultado operacional do produto, aumentar o número de apólices da carteira, reduzir a taxa de sinistralidade, entre outros. Para otimizar a variável ou variáveis escolhidas é normalmente necessário ajustar os preços propostos. Assim, com o auxílio do modelo de retenção, a companhia sabe qual a sensibilidade do cliente ao novo prémio e consegue ajustá-lo de forma a retê-lo. Por outro lado, se a companhia pretender alterar a composição do seu portfólio pode apresentar preços mais ajustados à taxa de retenção dos segmentos preferenciais. Para o processo de otimização ser iniciado é necessário que a companhia tenha modelos de risco, modelo de retenção e uma estratégia comercial. A Figura 8.1 apresenta o gráfico relativo a um exemplo do processo de otimização:

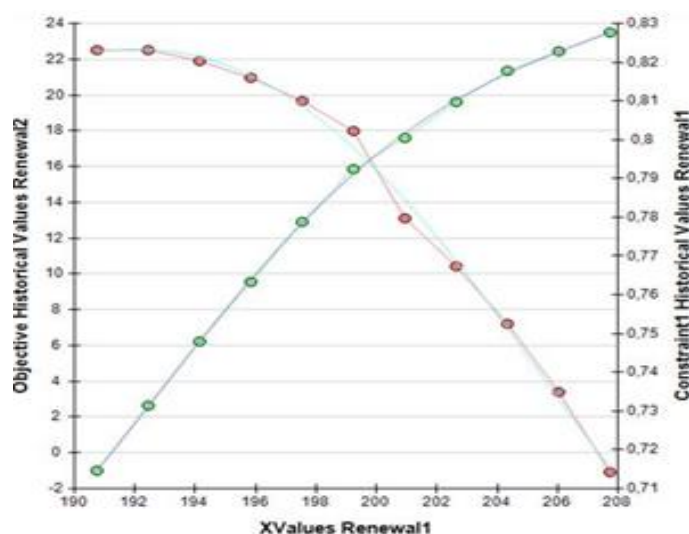


Figura 8.1 - Margem e Retenção

A curva a verde representa a margem de lucro. Verifica-se que o aumento do preço (eixo dos x's) faz com que se tenha maior margem (eixo dos y's à esquerda). Já a linha a vermelho corresponde à probabilidade estimada de retenção (eixo dos y's à direita). É possível ver que à medida que o preço aumenta, menor é a taxa de retenção. O ponto ótimo de otimização é dado pela intersecção da curva da margem e da retenção. Este processo é feito apólice a apólice, ou seja, cada cliente tem um determinado perfil de renovação. Assim sendo, e dependendo do objetivo comercial da companhia, é possível ajustar o preço que a companhia irá propor ao cliente com o intuito de atingir o seu objetivo.

Tal como em todas as áreas de negócio, um dos objetivos de uma companhia de seguros é obter lucro e, como tal, poderá ter de reduzir o seu número de apólices de forma a maximizar o seu resultado. Foi visto também ao longo deste trabalho que os consumidores são muito sensíveis ao preço e que a taxa de retenção é tanto maior quanto menores as alterações no preço.

Assim, a existência de um modelo de retenção em conjunto com o processo de otimização, permite tomar decisões sustentadas e distintas daquelas que tomaria em caso de não o ter.

Neste trabalho final de mestrado comparou-se também um modelo puramente estatístico com um modelo construído pela Generali. Os dois modelos foram construídos com a mesma base de dados e de seguida resume-se os pontos principais:

- O modelo construído apresenta as variáveis Bónus – Malus, Marca e valor do veículo enquanto que o modelo Generali não.
- O modelo construído não apresenta as variáveis zona geográfica e existência da cobertura de quebra isolada de vidros enquanto que o modelo da generali apresenta.
- Ambos os modelos apresentam as variáveis tipo de cliente, grupo de agentes, mês de renovação, existência da cobertura de danos próprios, prémio proposto, variação do prémio, distrito, idade do segurado, idade do veículo, número de anos de carta de condução, antiguidade da apólice, frequência de pagamento e retenção média do agente.
- Nenhum dos modelos apresenta as variáveis: franquias da cobertura de danos próprios, o prémio pago em 2015, o tipo de combustível, rácio entre o peso e a potência, e a idade do condutor.
- O modelo puramente estatístico apresenta um valor de AIC, AICc e BIC superior ao modelo utilizado pela seguradora Generali.
- Para os dois pontos de corte testados, o modelo utilizado pela Generali apresenta uma percentagem de observações bem colocadas superior ao modelo criado.
- O modelo Generali apresenta um valor de AUC superior ao modelo criado.

Considerando o resumo comparativo anterior, conclui-se que embora existam várias variáveis em comum, o modelo utilizado pela seguradora Generali encontra-se mais ajustado aos dados, traduzindo melhor a probabilidade de retenção de um perfil de apólice. Observa-se assim que a componente prática e a experiência de mercado são bons aliados num âmbito de identificar padrões e prever comportamentos humanos e que esta componente em conjunto com indicadores estatísticos e métodos teóricos de previsão produzem modelos mais robustos e ajustados à realidade.

Ambos os modelos apresentam uma discriminação fraca, pelo que se deve dar uma grande importância à sua revisão e sugere-se que esta seja feita com alguma periodicidade de

forma a captar novos comportamentos, tendências de mercado, regras de subscrição e legislação. Este foi um ajuste à realidade da empresa, pelo que pode variar ao longo dos anos. Sendo assim, aquando a revisão do mesmo, deve-se tentar incluir outras variáveis de modo a que se possa explicar melhor o comportamento dos clientes.

9. Referências

9.1. Referências Bibliográficas

David W. Hosmer, Jr., Lemeshow, S. , Sturdivant, R. (2003). *Applied Logistic Regression*, 3ª Edição

Manuel Sarmiento Batista ,António. (2015) Regressão Logística – Uma introdução ao modelo estatístico

Kleinbaum, David G., Klein, Mitchel (2010). Logistic Regression : A self-learning Text

Apontamentos da professora Teresa Alpuim para apoio à disciplina de Modelos Lineares do curso do Mestrado de Matemática Aplicada à Economia e Gestão.

Apontamentos da professora Maria Fernanda Oliveira para apoio à disciplina de Risco em Seguros Vida e não Vida do curso do Mestrado de Matemática Aplicada à Economia e Gestão.

Garraio, J. B. (Setembro de 2015). Modelação da Taxa de Anulação no Seguro Automóvel. FCUL-UL.

9.2. Referências Sitográficas

Grupo Generali. Disponível: <http://www.generali.com> [Acesso em: 2017/04]

Generali Portugal. Disponível : <https://ww6.generalip.pt/generali/pt/> [Acesso em: 2017/04]

Plataforma Digital “Intragen” disponível para trabalhadores da empresa Generali Portugal [Acesso em: 2017/04]

Curva Roc. Disponível: <https://www.medcalc.org/manual/roc-curves.php> [Acesso em: 2017/04]

Matriz de Confusão. Disponível: <http://developerdeveloper.blogspot.pt/2013/11/matriz-confusao.html> [Acesso em 2017/04]

10. Anexos

A. Tabela de Bónus-Malus

Escalão	Número de anos	% Desconto ou Agravamento	Número de Sinistros na última anuidade				
			Sem Sinistros	1 Sinistro	2 Sinistros	3 Sinistros	4 Sinistros
A	14	-53	A	E	I	K	O
B	13	-50	A	G	K	L	O
C	12	-50	B	H	K	M	O
D	11	-50	C	I	K	M	O
E	10	-48	D	I	K	M	O
F	9	-46	E	K	L	M	P
G	8	-44	F	K	M	N	P
H	7	-42	G	L	M	N	P
I	6	-40	H	L	M	O	P
J	5	-40	I	L	M	O	Q
K	4	-35	J	M	O	O	R
L	3	-30	K	N	O	Q	R
M	2	-20	L	O	O	R	S
N	1	-10	M	P	Q	S	T
O	0	0	N	Q	R	T	T
P		+10	O	R	S	T	T
Q		+20	P	S	T	T	T
R		+30	Q	T	T	T	T
S		+40	R	T	T	T	T
T		+75	S	T	T	T	T

B. Programa 1 – Calculo da taxa de Retenção

```
libname t 'C:\Users\henriquemarques\Documents\henrique\Tese\Base de dados';
```

```
data BaseDados;
set t. BaseDados;
run;
```

```
%macro retencao (variavel,tabela_final);
proc sql;
create table big_city as
select distinct &variavel., sum(Response) as num, sum(weight) as den, calculated num /
calculated den as taxa_retenção, count(numint) as apolices
from BaseDados
group by &variavel.; quit;
```

```
proc sql;
create table max as
select max(taxa_retenção) as max_tax_retencao
from big_city
```

```

;quit;

proc sql;
create table uniao as
select a.*, b.max_tax_retencao
from big_city as a left join max as b on a.taxa_retenção=b.max_tax_retencao
;quit;

proc sql;
create table big_city_uniao as
select &variavel., taxa_retenção, max(max_tax_retencao) as max_tax_retencao, apolices
from uniao ;quit;

proc sql;
create table &tabela_final. as
select &variavel., (taxa_retenção/max_tax_retencao) FORMAT=NLPCTI6. as
retenção_base_100 ,apolices
from big_city_uniao
order by &variavel. asc
;quit;

%mend;

%retencao(zona_geográfica,Regiontable);
%retencao(Tipo_Cliente,ClientRatingtable);
%retencao(Grupo_Agente,AgentGrouptable);
%retencao(Mês_Renovação,RenewalMonthtable);
%retencao(Tipo_Veículo_f,TypeVehicle_ftable);
%retencao(Cobertura_Vidros,WindscreenDedtable);
%retencao(Franquia_Casco,Franqu_Cascotable);
%retencao(Tem_DP,Has_Cascotable);
%retencao(Bonus_Malus,BMScale);
%retencao(PolicyAnualPremium_2015,PolicyAnualPremium_2015table);
%retencao(Proposed_Premium_,Proposed_Premium_table);
%retencao(VarPremium,VarPremium_PercBtable);
%retencao(Marca,Marca_ftable);
%retencao(Tipo_Combustivel,Fuel_ftable);
%retencao(Racio_Peso-Potencia_f,Weight_hp_ftable);
%retencao(Distrito_f,District_ftable);
%retencao(Valor_Veiculo,SI_OwnDamagetable);
%retencao(Idade_Condutor_f,AgeDriver_ftable);
%retencao(Idade_Segurado_f,Age_InsuredPerson_ftable);
%retencao(Idade_Veiculo_f,Age_Vehicle_ftable);
%retencao(Antiguidade_Carta_F,AGEDRIVERSLICENCE_Ftable);
%retencao(Antiguidade_Apolice_f,Policy_TimeOnBook_ftable);
%retencao(FreqPagamento_f,FreqPayment_ftable);
%retencao(Retencao_Media_Agente,Agent_retention_avgtable);

```


C. Programa 2 - Calculo do VIF

```
Libname t 'C:\Users\henriquemarques\Documents\Tese';
```

```
Data work.model;
```

```
Set t.BaseDados; run;
```

```
proc reg data=work.model;
```

```
model response=
```

```
Zona_Geografica Tipo_Cliente Grupo_Agentes Mês_Renovação Tipo_Veículo
```

```
Franquia_Cobertura_Vidros Franquia_Cobertura_DP
```

```
Has_DanosProprios Bonus_Malus_Apolice Percentagem_BM
```

```
PremioPago_2015 Premio_Proposto2016 VarPremium Marca Tipo_combustivel
```

```
Racio_Peso_Potência Distrito Capital_Seguro Idade_Condutor Idade_Segurado Idade_Veículo
```

```
Anos_Carta_Conducao Antiguidade_Apolice Forma_Pagamento Retencao_Agente
```

```
/vif ;run;
```

D. Programa 3 – Modelo de Regressão

```
Ods graphics on;
```

```
%_eg_conditional_dropds(WORK.PREDLOGREGPREDICTIONSQUERY__0001,  
WORK.temp);
```

```
/* -----
```

```
Sort data set Local:WORK.QUERY_FOR_TESTE_DUMMIES_SAS7BDAT
```

```
----- */
```

```
Proc Sql;
```

```
Create View work.BDModelo AS
```

```
Select *
```

```
from WORK.BaseDados
```

```
;
```

```
quit;
```

```
Title;
```

```
Title1 "Logistic Regression Results";
```

```
Footnote;
```

```
Footnote1 "Generated by the SAS System (&_SASSERVERNAME, &SYSSCPL) on
```

```
%TRIM(%QSYSFUNC(DATE(), NLDATE20.)) at %TRIM(%SYSFUNC(TIME()),
```

```
TIMEAMPM12.))";
```

```
proc logistic data=work. BDModelo
```

```
plots(only)=All
```

```
;
```

```
model Response (Event = '1')=Zona_Geografica Tipo_Cliente Dummy_Bronze
```

```
Dummy_Gold Dummy_Platinum Mes_Renovação Dummy_Ligeiros Dummy_Misto
```

```
Dummy_Caminheta Dummy_ttligeiro Dummy_ttmisto Dummy_ttcamineta Dummy_Comercial
```

```
Existencia_Cobertura_Vidros Franquia_DP_a) Franquia_DP_b) Franquia_DP_d)
```

```
Franquia_DP_e) Franquia_DP_g) Franquia_DP_h) Franquia_DP_i) Franquia_DP_j)
```

```
Franquia_DP_k) Franquia_DP_l) Franquia_DP_n) Tem_DP Percentagem_Bonus-Malus
```

```
_Premio_Proposto Variação_Prémio dummy_renault dummy_opel dummy_volkswagen
```

```
dummy_pegault dummy_citroen dummy_ford dummy_mercedes dummy_fiat dummy_toyota
```

```
dummy_bmw dummy_seat dummy_audi dummy_nissan dummy_mitsubishi dummy_honda
```

```

dummy_hyunday dummy_volvo dummy_smart dummy_skoda dummy_mazda dummy_rover
dummy_alfaromeo dummy_suzuki dummy_landrover dummy_kia dummy_lancia dummy_mini
summy_chevrolet dummy_iveco dummy_gamaalta dummy_erro_combustivel dummy_gasolina
dummy_gasnatural dummy_gasóleo Racio_Peso_Potência dummy_err dummy_aveiro
dummy_acores dummy_beja dummy_braga dummy_bragança dummy_castelobranco
dummy_coimbra dummy_faro dummy_guarda dummy_leira dummy_lisboa dummy_madeira
dummy_portalegre dummy_porto dummy_santarém dummy_setubal dummy_viana_do_Castelo
dummy_vilareal dummy_viseu Valor_Veículo Idade_Segurado Idade_Veículo
Anos_Carta_Contuação Antiguidade_Apolice Frequencia_Pagamento Retenção_Agente
/
    selection=Backward
    sls=0.05
    include=0
    rsquare
    link=Logit
;

    output out=WORK.output(Label="Logistic regression predictions and statistics for
work.testeddummy")
        predprobs=Individual;
run;
quit;

/* -----
End of task code.
----- */
run; quit;
%_eg_conditional_dropds(WORK.temp);
title; footnote;
ods graphics off;

```